AI4media
ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

D2.3

AI technologies and applications in media: State of Play, Foresight, and Research Directions

Project Title	AI4Media - A European Excellence Centre for Media, Society and Democracy
Contract No.	951911
Instrument	Research and Innovation Action
Thematic Priority	H2020-EU.2.1.1. - INDUSTRIAL LEADERSHIP - Leadership in enabling and industrial technologies - Information and Communication Technologies (ICT) / ICT-48-2020 - Towards a vibrant European network of AI excellence centres
Start of Project	1 September 2020
Duration	48 months

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu

Deliverable title	AI technologies and applications in media: State of Play, Foresight, and Research Directions
Deliverable number	D2.3
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	28 February 2022
Actual date of delivery	04 March 2022
Deliverable filename	AI4Media_D2.3_Roadmap_final
Nature of deliverable	Report
Dissemination level	Public
Number of pages	390
Work Package	WP2
Task(s)	T2.3
Partner responsible	CERTH
Author(s)	Filareti Tsalakanidou (CERTH)
Editor	Filareti Tsalakanidou (CERTH)
EC Project Officer	Evangelia Markidou

Abstract	This deliverable provides a detailed overview of the complex landscape of AI for the media industry. It analyses the current status of Artificial Intelligence (AI) technologies and applications for the media industry, highlights existing and future opportunities for AI to transform media workflows, assist media professionals and enhance the user experience in different industry sectors, offers useful examples of how AI technologies are expected to benefit the industry in the future, and discusses facilitators, challenges and risks for the wide adoption of AI by the media.
Keywords	Roadmap, Artificial Intelligence, Media, Creative industries, Entertainment, Online survey, Reinforcement learning, Evolutionary Learning, Scarce data, Transformers, Causal AI, Bioinspired learning, Quantum computing, Multimodal representation & retrieval, Media summarisation, Content creation, Affective analysis, Natural language processing, Content Moderation, Robust AI, AI fairness, Privacy-preserving AI, Explainable/ Interpretable AI, Datasets, Benchmarks, News production, Journalism, Disinformation, Robot journalism, Social media, Games, Music, Film/TV, Publishing, Social Sciences and Humanities, Citizen participation, Ethical and legal aspects, AI Ethics, AI bias, AI liability, AI transparency, AI regulation, AI Act, DSA Act, Environmental impact, Pandemic

Copyright

Contributors

NAME	ORGANISATION
Filareti Tsalakanidou	CERTH
Yiannis Kompatsiaris	CERTH
Symeon Papadopoulos	CERTH
Vasileios Mezaris	CERTH
Emmanouil Krasanakis	CERTH
Nikos Giatsoglou	CERTH
Evlampios Apostolidis	CERTH
Georgia Pantalona	CERTH
Lidia Dutkiewicz	KUL
Noémie Krack	KUL
Emine Ozge Yildirim	KUL
Nicu Sebe	UNITN
Cigdem Beyan	UNITN
Marco Formentini	UNITN
Kasim Sinan Yildirim	UNITN
Enver Sangineto	UNITN
Wei Wang	UNITN
Lorenzo Seidenari	UNIFI
Marco Bertini	UNIFI
Federico Becattini	UNIFI
Ioannis Pitas	AUTH
Vasileios Mygdalis	AUTH
Ioannis Patras	QMUL
Niki Foteinopoulou	QMUL
Ioannis Maniadis Metaxas	QMUL
Ioanna Ntinou	QMUL
James Oldfield	QMUL
Christos Tzelepis	QMUL
Georgios Zoumpourlis	QMUL
Mihai Gabriel Constantin	UPB
Mihai Dogariu	UPB
Anna Schjøtt Hansen	UvA
Tobias Blanke	UvA
Georgios N. Yannakakis	UM
Antonios Liapis	UM
Mara Graziani	HES-SO
Henning Muller	HES-SO
Adrian Popescu	CEA
Julien Tourille	CEA
Sven Becker	FhG-IAIS

NAME	ORGANISATION
Annerike Flint	FhG-IAIS
Laszlo Friedmann	FhG-IAIS
Andreas Steenpass	FhG-IAIS
Patrick Aichroth	FhG-IDMT
Thomas Köllmer	FhG-IDMT
Frederic Precioso	3IA-UCA
Lucile Sassateli	3IA-UCA
Giuseppe Amato	CNR
Alejandro Moreo Fernandez	CNR
Artur Garcia	BSC
Dario Garcia Gasulla	BSC
Daniel Gatica-Perez	IDIAP
Rémi Mignot	IRCAM
Axel Roebel	IRCAM
Hannes Fassold	JR
Werner Bailer	JR
Georg Thallinger	JR
Killian Levacher	IBM
Samuel C. Hoffman	IBM Yorktown Heights (NY)
Rasa Bocyte	NISV
Philo van Kemenade	NISV
Johan Oomen	NISV
Birgit Gray	DW
Mike Matton	VRT
Sabino Metta	RAI
Alberto Messina	RAI
David Melhart	MODL
Daniele Gravina	MODL
Ahmed Khalifa	MODL
Georgi Kostadinov	IMG
Chris Georgiev	IMG
Pavel Andreev	IMG
Ralitsa Golemanova	IMG
Julià Vicens	Eurecat – Centre Tecnològic de Catalunya
Mihnea Tufis	Eurecat – Centre Tecnològic de Catalunya
Cristian Consonni	Eurecat – Centre Tecnològic de Catalunya
Patricia Castillo	Eurecat – Centre Tecnològic de Catalunya
David Laniado	Eurecat – Centre Tecnològic de Catalunya

Peer Reviews

NAME	ORGANISATION
Andrea Esuli	CNR
Georg Thallinger	JR

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	25/11/2021	Filareti Tsalakanidou	Table of content, template for description of future AI trends
0.2	26/01/2022	Filareti Tsalakanidou	Updated version including partner contributions for sections 4-12
0.3	19/02/2022	Filareti Tsalakanidou	Updated version including the executive summary (section1), introduction (section 2), the survey analysis (section 4), the web-version of the roadmap (section 13), the conclusions (section 14), revisions to previous contributions in sections 4-12, and addition of new partner contributions (6.1, 6.3, 6.6, 7.3, 8.1, 8.2, 9.2, 9.3, 9.9)
0.4	17/02/2022	Filareti Tsalakanidou	Pre-final version ready for internal review
0.5	28/02/2022	Andrea Esuli (CNR), Georg Thallinger(JR)	Internal review
0.6	03/03/2022	Filareti Tsalakanidou	Updated version addressing internal review comments
0.7	04/03/2022	Filareti Tsalakanidou	Updated version (update of section 6.3, addition of content in section 9.6)
1.0	04/03/2022	Filareti Tsalakanidou	Final version ready for submission

Table of Abbreviations and Acronyms

Acronym	Meaning
2D	Two-dimensional
3D	Three-dimensional
5G	5^{th} Generation mobile network
AAAI	Association for the Advancement of Artificial Intelligence
AGI	Artificial General Intelligence
AI	Artificial Intelligence
AIIDE	Artificial Intelligence and Interactive Digital Entertainment
ALTAI	Assessment List on Trustworthy Artificial Intelligence
APD	Automated Process Discovery
API	Application Programming Interface
AR	Augmented Reality
ASR	Automatic Speech Recognition
AU	Action Unit
AVC	Advanced Video Coding
A/V	Audio/Video
BATX	Baidu-Alibaba-Tencent-Xiaomi
BERT	Bidirectional Encoder Representations from Transformers
BMK	Austrian Federal Ministry for Climate Protection, Environment, Energy, Mobility, Innovation and Technology
CAGR	Compound Annual Growth Rate
CAHAI	Ad hoc Committee on Artificial Intelligence
CBN	Causal Bayesian Networks
CCC	Computing Community Consortium
CEO	Chief Executive Officer
CG	Content Generation
CGI	Computer Generated Imagery
CLARIN	Common Language Resources and Technology Infrastructure
CMOS	Complementary Metal Oxide Semiconductor
CMU-MOSEI	CMU Multimodal Opinion Sentiment and Emotion Intensity dataset
CNN	Convolutional Neural Network
CO_2	Carbon Dioxide
CoE	Council of Europe
CompCI	Computational Causal Inference
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DH	Digital Humanities
DL	Deep Learning
DM	Data Management
DNI	Google's Digital News Innovation Fund

Acronym	Meaning
DNN	Deep Neural Network
DNS	Domain Name System
DOSN	Decentralised Online Social Network
DP	Differential Privacy
DPIA	Data Protection Impact Assessment
DPSGD	Differentially Private Stochastic Gradient Descent
DSA	Digital Services Act
DSM	Digital Surface Model
EaaS	Evaluation as a Service
EBU	European Broadcasting Union
EC	European Commission
ELSA	Ethical Legal and Societal Aspects
EP	European Parliament
E-RIS	European Research Infrastructure for Heritage Science
ESG	Environmental, Social, and Governance
EU	European Union
FJP	Facebook Journalism Initiative
FL	Federated Learning
fMRI	functional Magnetic Resonance Imaging
FoV	Field of View
FPS	First-Person Shooter game
GAFAM	Google-Amazon-Facebook-Apple-Microsoft
GAN	Generative Adversarial Network
GDP	Gross Domestic Product
GDPR	General Data Protection Regulation
GNI	Google's News Initiative
GNN	Graph Neural Network
GPT-3, GPT-4	Generative Pre-trained Transformer 3/4
GPU	Graphics Processing Unit
HD	High Definition
HE	Homomorphic Encryption
HLEG	High-Level Expert Group
HMD	Head-Mounted Display
HPC	High-Performance Computing
HR	Human Resources
HRIA	Human Rights Impact Assessment
ICT	Information Communication Technology
IEEE	Institute of Electrical and Electronics Engineers
IEMOCAP	Interactive Emotional Dyadic Motion Capture database
IL	Imitation Learning

Acronym	Meaning
ILSVRC	ImageNet Large Scale Visual Recognition Challenge
IMCO	Internal Market and Consumer Protection
IoT	Internet of Things
IP	Intellectual Property
IP	Internet Protocol
IPR	Intellectual Property Right
IT	Information Technology
JSON	JavaScript Object Notation
LIBE	EU Committee on Civil Liberties, Justice and Home Affairs
LIME	Local Interpretable Model-agnostic Explanations
MEG	MagnetoEncephaloGraphy
MEP	Member of the European Parliament
MIDI	Musical Instrument Digital Interface
MJP	Meta Journalism Project
ML	Machine Learning
MLP	Multi-Layer Perceptron
MPEG	Moving Picture Experts Group
MR	Mixed Reality
Mt	Metric tone
MXF	Material Exchange Format
NER	Named Entity Recognition
NeRF	Neural Radiance Fields
NGO	Non-Governmental Organisation
NISQ	Noisy Intermediate-Scale Quantum
NLP	Natural Language Processing
NPC	Non Playable Character
OA	Open Access
OCR	Optical Character Recognition
OSCE	Organisation for Security and Co-operation in Europe
p2p	peer-to-peer
PATE	Private Aggregation of Teacher Ensembles
PC	Personal Computer
PCG	Procedural Content Generation
PoC	Proof of Concept
PPAI	Privacy-Preserving AI
PR	Public Relations
PSM	Public Service Media
QA	Quality Assurance
RL	Reinforcement Learning
RNN	Recurrent Neural Networks

Acronym	Meaning
ROI	Return On Investment
RQC	Random Quantum Circuit
#SAIFE	Spotlight on Artificial Intelligence and Freedom of Expression
SEM	Structural Equation Modelling
SMPC	Secure Multiparty Computation
SSH	Social Sciences and Humanities
SSL	Self-Supervised Learning
TMT	Technology, Media & Telecommunications
TPU	Tensor Processing Unit
TRL	Technology Readiness Level
UDA	Unsupervised Domain Adaptation
UGC	User Generated Content
UK	United Kingdom
UMLS	Unified Medical Language System
UN	United Nations
URL	Uniform Resource Locator
US	United States
USD	United States Dollar
VAD	Valence-Arousal-Dominance
VAE	Variational AutoEncoder
VFX	Visual effects
VLOP	Very Large Online Platform
VQA	Visual Question Answering
VR	Virtual Reality
VSP	Video Sharing Platform
WP	Work Package
XAI	eXplainable Artificial Intelligence
XML	Extensible Markup Language
XR	Extended Reality

Index of Contents

Index of Tables

Index of Figures

1 Executive Summary

Deliverable D2.3 *"AI technologies and applications in media: State of Play, Foresight, and Research Directions"* aims to provide a ***detailed overview of the complex landscape of AI for the media industry***. It analyses the current status of Artificial Intelligence (AI) technologies and applications for the media industry, highlights existing and future opportunities for AI to transform media workflows, assist media professionals and enhance the user experience in different industry sectors, offers useful examples of how AI technologies are expected to benefit the industry in the future, and discusses facilitators, challenges and risks for the wide adoption of AI by the media.

For the analysis of the AI for Media landscape a multi-party, multi-dimensional and multi-disciplinary approach has been adopted, involving almost all AI4Media partners, external media or AI experts but also the AI research community and the community of media professionals at large. Three main tools have been used to accurately describe this landscape, including a ***multi-disciplinary state-of-the-art analysis*** involving AI experts, experts on social sciences, ethics and legal issues, as well as media industry practitioners; a ***public survey*** targeted at AI researchers/developers and media professionals; and a series of ***short white papers*** on the future of AI in the media industry that focus on different AI technologies and applications as well as on different media sectors, exploring how AI can positively disrupt the industry, offering new exciting opportunities and mitigating important risks.

Based on these tools, we provide a detailed analysis of the current state of play and future research trends with regard to media AI (short for "use of AI in media"), which is comprised of the following components.

State of the art analysis of AI technologies and applications for the media. Based on an extensive analysis of roadmaps, surveys, review papers and opinion articles focusing on the trends, benefits, and challenges of the use of AI, we provide a clear picture of the most transformative applications of AI in the media and entertainment industry. Our analysis identifies the following **AI applications that are already having or can have a significant impact** in most media industry sectors by addressing common needs and shared aspirations about the future: AI assistants, smart recommender systems, content personalisation, automatic content synthesis, multi-modal content search, multi-lingual translation, audience analysis, social media trend detection, forecasting tools, hyper-targeted advertisement and compliance with copyright standards. In addition, we identify a list of ***AI technologies that hold the greatest potential*** to realise the media's vision for AI. Finally, we identify ***AI challenges*** with regard to AI's trustworthiness, including AI explainability, robustness against adversarial attacks, AI fairness, and data privacy issues.

Discussion of social, economic and ethical implications of AI. Complementing the previous state of the art analysis, which highlights AI's potential for the media industry from a technology and practical application point of view, this analysis dives into the ***social and ethical implications of AI***, offering the point of view of social scientists, ethics experts and legal scholars, based on an extensive literature review of both industry reports and scholarly journal articles. The most prevalent societal concerns and risks regarding AI for media are identified, which include: biases and discrimination; media (in)dependence and commercialisation; increased inequality to access to AI; labour displacement, monitoring and profession transformations; privacy, transparency, accountability and liability; manipulation and misinformation as an institutional threat; and environmental impacts of AI. In addition, we identify what practices will be important to ***counteract potential negative societal impacts*** of media AI.

Summary of EU policy initiatives and their impact on future AI research for the media. We provide an overview of the ***most important EU policy initiatives on AI***, focusing on initiatives having a ***clear focus on the media industry***. We discuss both policy initiatives (non-binding provisions) and regulatory initiatives (leading to the adoption of binding legal provisions), including the European Parliament resolution on AI in education, culture and the audiovisual sector; the Digital Services Act proposal (DSA Act); the Proposal for a Regulation laying down harmonised rules on artificial intelligence (AI Act); the Code of Practice on disinformation; and the Proposal on transparency and targeting of political advertising. The analysis of these initiatives shows that the use of AI media applications is on the verge of being specifically regulated in legal instruments. The common approach chosen to deal with them is principally transparency requirements and obligations and then, at a lesser level, empowerment of users.

Analysis of survey results, providing insights about the hopes, aspirations and concerns of the AI research community and the community of media practitioners. Two online surveys have been launched to help us collect the opinions of AI researchers and media professionals on the use of AI for media: i) a public survey aiming to collect the opinions of the AI research community and media industry professionals with regard to the benefits, risks, technological trends, challenges and ethics of AI use in the media industry ; and b) a small-scale internal survey addressed to AI4Media partners, aiming to collect their opinions on the benefits and risks of media AI for the society and democracy.

150 responses from AI researchers and media professionals from 26 countries in Europe and beyond were collected as part of the first survey. The main findings include:

- Automatic content analysis & creation, multi-lingual NLP, learning with limited data, explainable AI and fair AI are considered the AI technologies & applications with the biggest potential to solve existing problems of the media industry.
- Automation & optimisation of routine tasks, personalisation of content and services, and increased productivity & operational efficiency are considered to be the most important benefits of AI for the media industry while unethical use of AI, AI bias, AI's lack of explainability, high expectations & low return, and failure of AI in a critical mission are some of the biggest risks.

- With regard to challenges for the adoption of AI, AI researchers focus on the lack of data to train AI algorithms while the media industry highlights integration of AI in business operations and processes and lack of AI skills in the industry.
- Most organisations (58% of respondents) do not have a clear AI strategy in place; furthermore, many do not have ethical frameworks to manage relevant risks (17% of respondents) or are unaware of the existence of such frameworks (19% of respondents).
- There is a clear preference of respondents towards open AI solutions and collaborations (e.g. open source development tools, ready to use components from AI repositories, and co-development with research partners) that will lead to the adoption of AI that they can trust.
- More guidance by policy makers with regard to the implementation of AI ethics is required on topics such as data privacy; IP issues; AI bias; automated content creation; development of trustworthy and explainable AI; and regulation impact.

31 responses were received as part of the second survey. The main findings include:

- The most promising applications of AI that can benefit society and democracy are counteracting disinformation, enabling equitable access for all, and providing tools to hold power accountable.
- The risks that worry respondents the most with regard to their potentially detrimental impact on society and democracy are unauthorised profiling & monitoring of citizens, use of AI to spread online disinformation & fuel polarisation on online discussion, and AI bias.
- The policies that governments should mainly adopt to ensure respect of fundamental rights by media companies are oversight by independent authorities and codes of practices. The least preferred choice is regulation. With regard to the policies that media organisations should adopt, the most popular ones include revealing information on their algorithms and data, publishing yearly reports on issues with ethics repercussions, and empowering users.

Analysis and future outlook of main AI technology & research trends for the media sector. Based on the results of the state of the art analysis, we highlight the potential of specific AI technologies to benefit the media industry. These include ***reinforcement learning, evolutionary learning, learning with scarce data, transformers, causal AI, AI at the edge, bioinspired learning and quantum computing for AI learning***. For each technology, a short ***white paper*** has been prepared aiming to offer a clear overview of the current status of the technology, drivers and challenges for its development and adoption, and future outlook. These white papers also include vignettes, i.e. short stories with media practitioners or users of media services as the main characters, aiming to showcase how AI innovations could help the media industry in practice. The analysis highlights the significant transformation that some of these technologies can bring in the short-term (reinforcement learning, learning with scarce data, transformers) but also in the long-term (bioinspired learning, quantum computing).

Analysis and future outlook of main AI applications for the media sector. Based on the results of the state of the art analysis, we highlight the potential of specific AI applications to benefit the media industry. These include ***multimodal knowledge representation and retrieval (content indexing and search), media summarisation, automatic content creation, affective analysis NLP-enabled applications like conversational agents, and content moderation***. Similarly to above, a short white paper has been prepared for each application aiming to offer a clear overview of the current status of the technology, drivers and challenges for its development and adoption, and future outlook. The analysis sheds light to the amazing potential of these technologies to provide effective solutions to long-standing problems of the industry.

Overview of the future of AI in different media sectors. We present a collection of white papers, focusing on the deployment of AI on different media industry sectors, including ***news, social media, film/TV, games, music and publishing***. We also explore the use of AI to address critical global phenomena such as ***disinformation*** and to enhance ***online political debate***. Finally, we explore how AI can help the study of media itself in the form of ***AI-enabled social science tools***. Following the same format as before, these papers offer an in-depth look on the current status of each sector with regard to AI adoption, the most impactful AI applications, the main challenges encountered, and future outlook. The analysis reveals a complex but colourful media landscape in the future, where AI is used to provide solutions to major media industry problems.

Analysis of future trends for trustworthy AI. We present four white papers (following the same format as above) focusing on different aspects of trustworthy AI, namely ***AI robustness, AI explainability, AI fairness, and AI privacy***, with a focus to media sector applications. The analysis explains existing trustworthy AI limitations and potential negative impacts and highlights how AI cannot be adopted at a scale in the media industry unless more emphasis and resources is put on this issue.

Overview of AI datasets and benchmarks. We analyse existing ***AI datasets and benchmark competitions***, discussing current status, research challenges and future outlook, while also providing insights on the ***ethical and legal aspects*** of availability of quality data for AI research.

Overview on AI democratisation. We discuss issues related to AI democratisation, focusing on ***open repositories for AI algorithms and data*** and research in the direction of ***integrated intelligence***, i.e. AI modules that could be easily integrated in other applications to provide AI-enabled functionalities. The analysis highlights the importance of openness (for data, storage and computational resources, algorithms, and AI talent) in the development of AI systems, focusing on modular toolboxes and open source software as the future of open AI repositories.

Exploration of external forces that could shape the future of AI for media. We discuss the forces that could shape the future of the use of AI in the media sector, focusing on ***legislation/***

regulation, the ***pandemic*** and its impact, and the ***climate crisis***. With regard to legislation, we predict that AI media applications will become more and more regulated in the next few years, affecting AI research and media organisations in various ways. With regard to the pandemic, we argue that the pandemic created exciting new opportunities for the growth of the media and entertainment industry and the expansion of their services and audience, while at the same time brought significant operational and creative challenges. These trends consolidate the position of AI as a transformative power in this industry, capable of revolutionising how operations run and how content is created, delivered and consumed. Finally, with regard to the environmental crisis, we encourage the media industry and AI researchers to be more proactive in making sure that AI technologies take environmental considerations into account. An understanding of the costs of building and deploying AI models not only financially but also in terms of environmental impact must be developed across the board and the mentality that more data or bigger models are always better should be re-examined.

The current report provides a detailed analysis of the current state of play and future research trends with regard to media AI, covering all aspects mentioned above and resulting in a comprehensive document of 390 pages. In addition to the full report and in order to help potential readers focus on the aspects that are most relevant to their own expertise and interests, we have also created **a web version of the report**[1] as part of the AI4Media website. This allows readers to easily go through the contents of this report though a user-friendly visual interface and only read/download the subsections covering topics that are of interest to them.

[1] Web version of the AI4Media Roadmap on AI technologies and applications for the Media Industry: https://www.ai4media.eu/roadmap-ai-for-media/

2 Introduction

The world is changing. Following a series of breakthroughs in the field of ***Artificial Intelligence (AI)***, new technologies are emerging which are ushering a wave of revolutionary innovations in nearly all aspects of business and society; from transportation[2] to finance[3], the fight against climate change[4], the media industry[5], journalism[6], and politics[7].

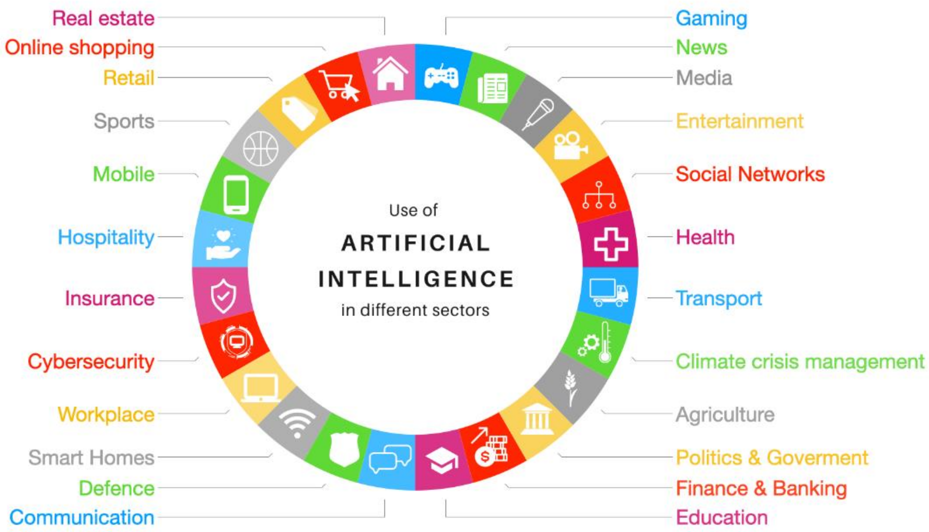


Figure 1: Industries and areas of life to be disrupted by Artificial Intelligence[8].

In all facets of economic and social life, AI is disrupting existing practices and creates opportunities for accelerated technological progress and global economic growth and development, promising to make our professional and personal lives easier through increased

---

[2] European Parliamentary Research Service, Artificial intelligence in transport - Current and future developments, opportunities and challenges (2019): https://www.europarl.europa.eu/RegData/etudes/BRIE/2019/635609/EPRS_BRI(2019)635609_EN.pdf

[3] OECD Business and Finance Outlook 2021 - AI in Business and Finance (2021): https://www.oecd-ilibrary.org/sites/39b6299a-en/index.html?itemId=/content/component/39b6299a-en#:~:text=Artificial%20intelligence%20(AI)%20is%20increasingly,and%20enhance%20the%20product%20offerings.

[4] K. Hao, MIT Technology Review (2019): Here are 10 ways AI could help fight climate change: https://www.technologyreview.com/2019/06/20/134864/ai-climate-change-machine-learning/

[5] C. Dilmegani, Top 17 AI Trends/Applications in Media & Entertainment 2022 (2022): https://research.aimultiple.com/ai-media/

[6] C. Underwood, Automated Journalism – AI Applications at New York Times, Reuters, and Other Media Giants (2019): https://emerj.com/ai-sector-overviews/automated-journalism-applications/

[7] M. Rissehttps, Artificial Intelligence and the Past, Present, and Future of Democracy (2021): https://carrcenter.hks.harvard.edu/files/cchr/files/ai-and-democracy

[8] Figure inspiration from https://www.oneragtime.com/24-industries-disrupted-by-ai-infographic/

automation[9], to provide solutions for and achieve breakthroughs in major world problems like poverty[10], climate crisis or cancer[11], to ensure equitable access for all[12], to increase productivity, innovation and creativity[13], to empower communities and strengthen democracy[14], and to create a safer and better world for all.

Although the potential of AI seems unlimited, it also comes with a considerable amount of ethical challenges and risks. While it can generate value for business and prosperity for society, it also gives rise to a host of serious consequences, some of them visible (e.g. violation of personal privacy by unauthorised user profiling, discrimination against underrepresented groups of citizens, accidents caused by autonomous AI systems, manipulation of public opinion through disinformation, violation of fundamental rights like the freedom of expression through questionable moderation choices, just to name of few) but also many other that we do not fully grasp yet[15]. In order to safely and responsibly enjoy the benefits of AI, we should at the same time be ready to mitigate its various risks. This necessitates a greater focus on issues of trust, ethics and accountability, besides the pursuit of technological progress and economic growth.

Such a human-centric, trustworthy and ethical brand of AI is particularly relevant to the media sector. Digital media permeates most aspects of human and social activity and is intertwined with information exchange and knowledge transfer. Machine vision and visual content understanding were some of the first fields to exhibit significant breakthroughs in the evolution of AI, including advances in audio/music analysis and generation, text and language analysis, and modelling of social trends. The media market is already benefiting from AI-based support across the value chain: for media newsgathering, production, distribution, and delivery as well as audience analysis. This includes a range of tools and services for processes such as information analysis, content creation, media editing, content optimisation, audience preference analysis, and recommender systems[16].

AI technologies are expected to disrupt the media and entertainment industry through advances in content synthesis, analysis, and distribution, but also by offering new deeper in-sights into

[9] J. Marsh, The Intelligence Revolution: 4 ways that AI makes life easier (2021): https://datafloq.com/read/10-ways-automation-makes-life-easier-everyone/

[10] J. Bennington-Castro, AI Is a Game-Changer in the Fight Against Hunger and Poverty. Here's Why (2019): https://www.nbcnews.com/mach/tech/ai-game-changer-fight-against-hunger-poverty-here-s-why-ncna774696

[11] C. Luchini, A. Pea, and A. Scarpa. Artificial intelligence in oncology: current applications and future perspectives. Br J Cancer 126, 4–9 (2022): https://doi.org/10.1038/s41416-021-01633-1

[12] C. Martinez, Artificial Intelligence and Accessibility: Examples of a Technology that Serves People with Disabilities (2021): https://www.inclusivecitymaker.com/artificial-intelligence-accessibility-examples-technology-serves-people-disabilities/

[13] B. Dickson, The Artist in the Machine: The bigger picture of AI and creativity (2020): https://bdtechtalks.com/2020/04/22/artist-in-the-machine-ai-creativity/

[14] K. Johnson, How AI can empower communities and strengthen democracy (2020): https://venturebeat.com/2020/07/04/how-ai-can-empower-communities-and-strengthen-democracy/

[15] B. Cheatham, K. Javanmardian, and H. Samandari, Confronting the risks of artificial intelligence (2019): https://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/confronting-the-risks-of-artificial-intelligence

[16] F. Tsalakanidou et al, “The AI4Media project: Use of Next-generation Artificial Intelligence Technologies for Media Sector Applications”, Proc. AAAI 2021: https://doi.org/10.1007/978-3-030-79150-6_7

the complex and rapidly evolving social processes that unfold online and offline by sensing citizen activities, interests and opinions. AI technology could help shape the media experience for users by enabling new ways of being informed, being entertained, being creative, interacting with content, communicating with other people all over the world, etc.

It will also transform the existing workflows of the media industry[17]:

- by automating routine or tedious processes (from content creation and content search to automatic analysis of legal documents to ensure compliance with copyright standards);
- by developing AI assistants that can support media professionals in their daily tasks (e.g. when writing an article or creating visual assets for a new game);
- by improving audience analysis and user profiling to offer better content and services to users;
- by offering advanced forecasting capabilities and other decision support tools that will facilitate better short-term and long-term business decisions by management and staff.

AI technologies can support the relationship between media providers and their audiences, helping to align with the needs of media users and citizens. The use of AI can also cut down operating costs and ultimately free up resources that can be directed to support work of better quality and increased creativity. Moreover, AI can create opportunities for the better realisation of public values, such as media diversity, freedom of expression, and inclusiveness. Finally, AI can help legacy media to be more competitive in a digital marketplace that is currently dominated by large platforms.

This report aims to provide an accurate overview of the current landscape with regard to the use of AI on the media sector. Our aim is to:

- ***collect the opinions*** of members of the AI research community and the media industry, with regard to the most important AI trends, benefits, risks, and challenges for the use of AI in the media sector, also collecting their insights on handling AI ethics;
- shed light on ***AI research and technology trends***, exploring the potential of emerging technologies for the media industry;
- explore techniques for ***trustworthy media AI***, focusing on issues of robustness, fairness, explainability and privacy;
- look into issues of ***data collection and algorithm benchmarking*** for media-related tasks, also examining relevant legal aspects;
- investigate how ***open AI repositories and integrated intelligence*** can accelerate AI development, facilitate adoption by the media industry, and ensure democratised access to AI offerings by small and large organisations and individuals;
- explore the wide range of ***AI-enabled applications for the media industry***;

[17] Dataiku, Ink., The AI Disruption in Media & Entertainment (2020): https://content.dataiku.com/ai-media-entertainment/ai-media-entertainment

- predict how AI can ***transform the different sectors of the media industry***, presenting informative scenarios of possible use and analysing opportunities and challenges;
- examine ***ethical, societal, economic, environmental, etc. concerns & risks*** for the use of AI in the media sector, and also presenting possible mitigation measures;
- summarise relevant ***EU policy initiatives and proposed regulations*** and discuss their impact on future AI research for the media industry; and
- explore the ***impact of*** external forces like ***legislation initiatives, the pandemic, and the climate crisis*** on the development & deployment of AI in the media sector.

In the following subsections of this Introduction, we try to explain what AI is and why it is important for the media sector. We also describe the methodology and tools used for the preparation of this report and present the structure of the rest of this document.

2.1 What is Artificial Intelligence?

The previous paragraphs offered some first insights on how AI is disrupting every industry but also every aspect of our life, already making considerable impact in various areas. Interestingly though, despite AI's rapid technological advancements and accelerated adoption by business and society, an exact definition of Artificial Intelligence is still elusive, 72 years after Alan Turing asked the question "*Can machines think*?" in his seminal paper "Computing Machinery and Intelligence"[18], which ignited the first discussions around AI.

In this report, we use the 1955 definition by John McCarthy, considered to be the father of AI, who explained **Artificial Intelligence** as "***the science of making machines do things that would require intelligence if done by people***"[19]. This inclusive definition allows us to understand AI as "an umbrella term comprising several techniques"[20] and encompassing a broad range of technologies, as shown in Figure 2.

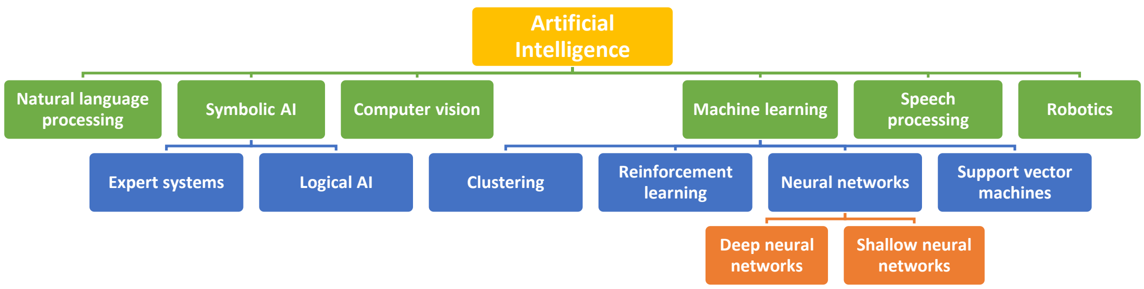


Figure 2: Branches of Artificial Intelligence[21].

[18] A. M. Turing, Computing Machinery and Intelligence, Mind, Volume LIX, No 236, p. 433–460 (1950): https://doi.org/10.1093/mind/LIX.236.433
[19] J. McCarthy, M. L. Minsky, N. Rochester, and C.E. Shannon, A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence (1955): http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf
[20] Broadband Commission, Reimagining Global Health through Artificial Intelligence: The Roadmap to AI Maturity (2020): https://broadbandcommission.org/Documents/working-groups/AIinHealth_Report.pdf
[21] Diagram adapted from: https://broadbandcommission.org/Documents/working-groups/AIinHealth_Report.pdf

Three terms closely related to AI are intelligence, machine learning (ML) and deep learning (DL). In the following, we provide the definitions of these terms as proposed by Stanford's Institute for Human-Centered AI[22]:

Intelligence is the "ability to learn and perform suitable techniques to solve problems and achieve goals, appropriate to the context in an uncertain, ever-varying world. A fully pre-programmed factory robot is flexible, accurate, and consistent but not intelligent".

Machine learning is the "part of AI studying how computer agents can improve their perception, knowledge, thinking, or actions based on experience or data. For this, ML draws from computer science, statistics, psychology, neuroscience, economics and control theory".

Deep Learning is the "use of large multi-layer (artificial) neural networks that compute with continuous (real number) representations, a little like the hierarchically organised neurons in human brains. It is currently the most successful ML approach, with better generalisation from small data and better scaling to big data and compute budgets".

Although, AI and ML are relatively new fields of research they are "built upon a long history of philosophical and scientific developments, including areas such as philosophy, ethics, logic, mathematics, and physics"[23]. Other disciplines that most recently have been contributing to AI include data analytics, statistical modelling, cybersecurity and encryption, while hardware devices and networks, underlying logics and principles, and user interface and user experience design are also important parts of the AI equation[23]. At the same time, a variety of technologies comprise AI, including natural language processing, computer vision, speech processing, machine learning and its branches like neural networks and reinforcement learning, expert systems, robotics, etc. as shown Figure 2.

A nice summary of AI's foundational principles, contributing disciplines/areas, and AI technologies can be found on the white paper on the "Spectrum of Artificial Intelligence" by the Future of Privacy Forum[23]. Taken from this paper, Figure 3 offers an informative visualisation of the ***spectrum of AI*** and its variety and complexity, showing the main types of AI and their relationships, characteristic examples of current AI applications in business and private life, and an overview of the main disciplines and scientific areas that AI is built upon or informed by.

[22] These definitions for Intelligence, Machine Learning and Deep Learning were proposed by Stanford's Institute for Human-Centered AI: https://hai.stanford.edu/sites/default/files/2020-09/AI-Definitions-HAI.pdf

[23] B. Leong and S.R. Jordan, The Spectrum of AI: Companion to the FPF AI Infographic (2021): https://fpf.org/wp-content/uploads/2021/08/FPF-AIEcosystem-Report-FINAL-Digital.pdf

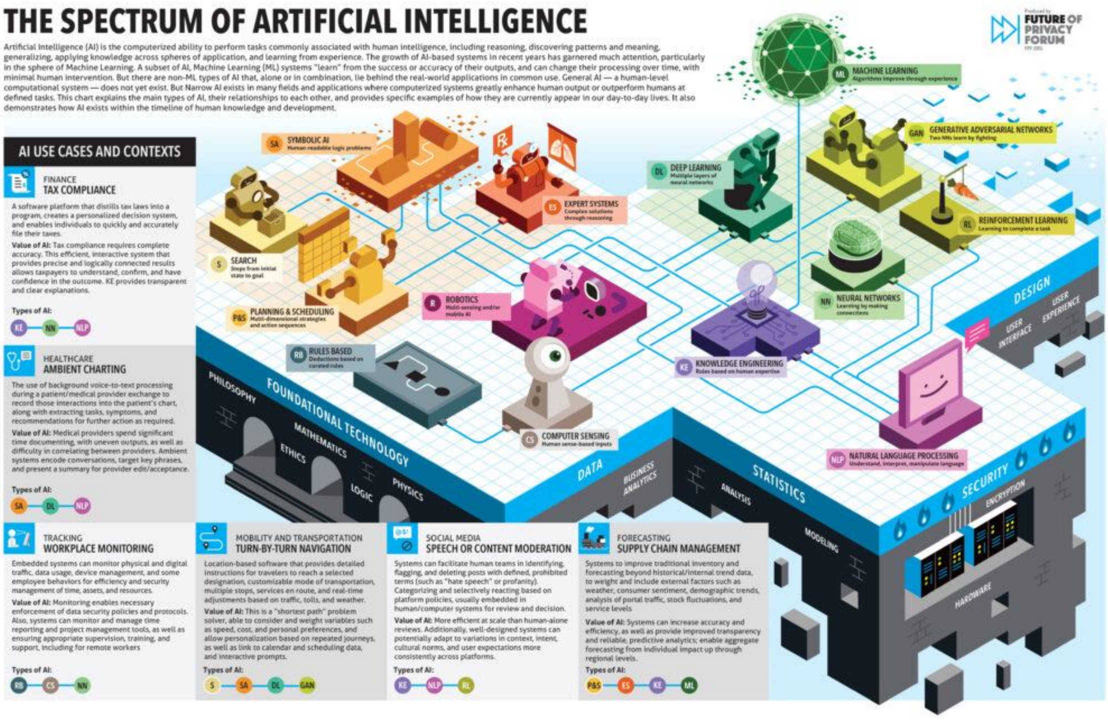


Figure 3: The spectrum of Artificial Intelligence as identified by the Future of Privacy Forum[24].

An ***overview of the history of AI***, presenting some of the most important milestones is presented in Figure 4. In 1950s, the Turing test is proposed by Alan Turing while in 1955 John McCarthy offers a first definition of AI. In 1964, a pioneering chatbot named ELIZA is developed at MIT. In 1997, IBM's Deep Blue beats Garry Kasparov in a chess competition. In 2002, iRobot mass produces Roomba, an autonomous robotic vacuum cleaner. In 2006, deep neural networks start becoming popular thanks among others to the work of Geoffrey Hinton[25,26]. This is an important milestone that drives the exponential growth of deep learning in the years after. Although, there were competitions on self-driving cars starting from the nineties, in 2009, Google builds the first self-driving car. In 2011, Apple's Siri and IBM's Watson chatbots are designed, with Watson winning against humans in Jeopardy. In 2012, AlexNet a convolutional neural network trained on a GPU significantly outperforms its competitors, achieving a jump in accuracy by more than 10%. In 2014, a chatbot called EUGENE passes the Turing test. In 2015, the ImageNet challenge declares that, for the first time, computers can more accurately identify objects in images than humans and in 2016, Google trains a neural network with 10 million unlabelled images that could

---

[24] Image source: Future of Privacy Forum, The Spectrum of Artificial Intelligence - An Infographic Tool (2021): https://fpf.org/blog/the-spectrum-of-artificial-intelligence-an-infographic-tool/

[25] Hinton, G. E., Osindero, S., & Teh, Y.-W. (2006). A fast learning algorithm for deep belief nets. Neural Computation, 18(7), 1527–1554.

[26] Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504–507.

accurately detect pictures of cats. In 2017, Google's AlphaGO was the first program to defeat a professional human Go player.

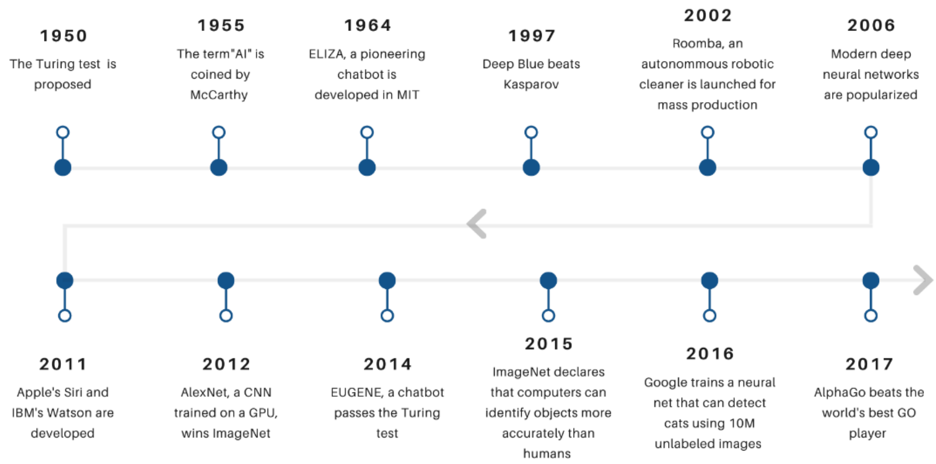


Figure 4: A history of Artificial Intelligence[27].

While AI has been slowly penetrating every industry in the last decade, its adoption has skyrocketed in the last couple of years, mainly driven by the Covid-19 pandemic, which brought AI and data analytics' potential to the forefront, accelerating the digital transformation of businesses with a rate never seen before[28]. According to a PwC study[29], 86% percent of companies surveyed said that AI was becoming a "mainstream technology" at their company in 2021. At the same time, an AI Journal study[30] reveals that 72% of business leaders feel positive about AI's role in the future, believing that it will make business processes more efficient (74%).

In this global environment, AI is "*set to be the key source of transformation, disruption and competitive advantage in today's fast changing economy*", expected to boost the global GDP by 14% by 2030, contributing $ 15.7 trillion to the global economy[31]. These gains will come from productivity improvements in the workplace driven by automation and AI assistants, personalisation of products and services for a better user experience, and improvement of the quality of offered services and products that will significantly increase consumer demand.

[27] The timeline in this figure was inspired by a) https://connectjaya.com/ai-timeline/, b) https://digitalwellbeing.org/artificial-intelligence-timeline-infographic-from-eliza-to-tay-and-beyond/, and c) https://blog.hurree.co/blog/the-history-of-artificial-intelligence-infographic

[28] J. McKendrick, Harvard Business Review, AI Adoption Skyrocketed Over the Last 18 Months (2021): https://hbr.org/2021/09/ai-adoption-skyrocketed-over-the-last-18-months

[29] PwC, AI Predictions 2021 (2021): https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions.html

[30] AI Journal, AI in a post-COVID-19 world (2021): https://aijourn.com/report/ai-in-a-post-covid-19-world/

[31] PwC, Sizing the prize - What's the real value of AI for your business and how can you capitalise?: https://www.pwc.com/gx/en/issues/data-and-analytics/publications/artificial-intelligence-study.html

In the next sub-section we briefly explore what all this means for the media & entertainment industry, highlighting how AI can transform workflows, decision-making, creation and delivery of content, as well as user experience.

2.2 Why Artificial Intelligence for the Media & Entertainment?

The media & entertainment industry (news, film/TV, music, games, social media, advertisement, publishing etc.) is already benefitting from AI advancements that can significantly facilitate, enhance or transform important tasks across the media industry value chain, including but not limited to: automation of existing tedious workflows; automatic content enhancement and creation; personalisation of content and services via enhanced user profiling; improved content recommendations; accurate audience analysis for enhanced audience targeting, content/ services development and increased advertisement revenue, at the global but also at the local level; improved accessibility to content thanks, for example, to automatic language translation; accurate forecasting about different businesses aspects; and more efficient decision making in general.

In the following, we briefly summarise the areas in which there is the greatest opportunity for AI to have a significant impact in the media industry by offering solutions in some of the most pressing problems of the industry[32]. As we can see, the transformative role of AI has already started to manifest in many of these areas, with important breakthroughs over the last few years in some cases. A more in-depth analysis of this landscape is offered in section 3 of this report.

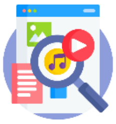
Automation of tedious tasks and AI assistants for increased productivity. Media workflows often include tedious or boring tasks, requiring a lot of resources. Some examples include searching large audio-visual archives or the Internet to locate information that can help a fact-checker verify the validity of some statement, analysing large volumes of documents for investigative journalism, producing subtitles or voice dubbing in different languages, producing content summaries, moderating content, organising A/B tests for different product parameters, clarifying complex IPR, etc. AI can help media professionals do their job more efficiently either by completely automating some tasks (e.g. content labelling or multi-lingual translation) or supporting professionals in more creative tasks (e.g. by offering automated suggestions, editing or enhancing content, answering questions, offering predictions about user engagement with content, etc.).

Content & services personalisation. With tons of content and a large variety of services available out there for the audience to enjoy (from news to films, music, games, books, graphics, etc.), media companies are in a constant battle against competitors for the audience's interest, trying to minimise churn rates, maximise user engagement with their content and attract new users. In this race, content personalisation seems to be the winning horse, with more and more media companies investing large amounts

[32] The icons used in this and the next page are from flaticon.com

of money to personalise their content and services and thus satisfy each customer's unique preferences, experiences, needs and moods. Elaborated profiling based on the continuous collection and analysis of user preferences, behaviours, and actions is already widely used in many media sectors (e.g. gaming industry, social media, advertisement, streaming services, etc.), however the trend is moving towards more elaborated approaches that also consider what happens to the user or in the world at the moment. Personalisation encompasses content suggestion, content presentation, interaction with content or personalisation of content itself (e.g. personalised movie trailers). It also means providing content to users where they are and when they want it.

Automated content creation. One of the biggest issues of the media industry is the ever-growing demand for new content. During recent years, AI advances, especially in the areas of generative AI, computer vision and natural language processing, have offered several solutions in this direction by enabling the automatic synthesis of new content based on the use of existing text, video, audio files, or images. The applications are already numerous: procedural content creation for games, deepfakes for the film industry, robot journalism, automated summaries for books and films, creation of new music, generation of script and visuals for advertisement, etc. Automated content creation can increase productivity and creativity in the media industry but also provide new ways of creativity for the general public.



Content indexing and search. The sheer volume of content generated everyday by the media industry nowadays is unprecedented: news items, films, books, music and songs, advertisements, social media posts, reviews, user generated videos, etc. This creates considerable challenges when it comes to efficient content labelling, search & retrieval processes, especially in the case of video and audio, and stands in the way of efficient content monetisation. AI promises to lift these obstacles by exploiting advanced video, audio and natural language analysis for content (e.g. detection and recognition of faces, voices, objects, places, dates, context etc.) that will enable automatic content labelling and will move beyond simple text queries to support visual search or complex voiced questions. This will allow fast and efficient search on large audio-visual archives as well as on the Internet for both media professionals and users aiming to find content that fits specific criteria (e.g. belongs to specific era, shows a specific person, involves a specific type of event – from earthquakes to music concerts-, includes specific human activities etc.). It will improve automatic content recommendation by offering suggestions that match user interests with the actual 'content' of the content, and it will allow media companies to more effectively exploit existing content and profit from it.

Audience analysis. Understanding what the audience wants or needs or how the audience feels is perhaps the number one priority of the media industry. AI and data science have already transformed audience analysis by allowing large-scale collection and analysis of user behaviours, emotions, actions, interactions with content, providing unprecedented insights to audience needs, wants and moods, allowing media companies to more effectively target different audiences and monetise their content. In addition, trend detection allows media companies to react in real-time to what is happening around the world and adapt accordingly.

Forecasting. Predictive analytics can facilitate short-term decisions but also the design of long-term strategies. Accurate predictions with regard to, among others, content engagement and monetisation, user behaviours, sales or churn rates, ad revenue, industry trends etc. can decisively improve decision-making mechanisms in the media industry, allowing for a timely reaction and efficient adaptation to a fast-changing reality.

While the benefits of AI for the media industry are many and important, they do not come without ***significant challenges and risks***. The first and most important is ***the risk posed to user privacy*** by the large-scale user monitoring and profiling mechanisms used by the media industry in order to offer increased personalisation and achieve better user targeting. Equally, disturbing are the phenomena of ***AI bias and discrimination*** against specific groups of people, including racial bias, gender bias, etc. For example, recommendation engines may discriminate against women when trained with film reviews that are mainly contributed by men while NLP models may introduce bias against underrepresented groups.

Another significant concern is that of ***lack of AI explainability***, with AI systems currently being black boxes that are not able to explain how they reached a decision, e.g. recommending specific content or predicting an outcome. More transparency is required about how AI tools work in order for media professionals to trust them.

There is also a growing concern regarding ***manipulation of content and misinformation***, making media organisations fear about the negative impacts of the growing amounts of misinformation to the public's trust in the media but also to the freedom of expression. While media companies become more deeply embedded into the platform economy mainly driven by AI there is also concern about the ***commercialisation of media organisations*** and how that affects their independence or their social responsibility. And finally, there is a growing concern among media professionals about how the increased automation of media workflows enabled by AI may lead to ***loss of human jobs*** or negatively affect creativity.

This report attempts to analyse this complex landscape of AI in the media sector, highlighting opportunities for growth and transformation but also discussing relevant risks and mitigation measures.

2.3 Methodology adopted

To develop this first version of the roadmap three tools have been mainly used: state of the art analysis; online surveys; and development of mini white papers with a look into the future (Figure 5).

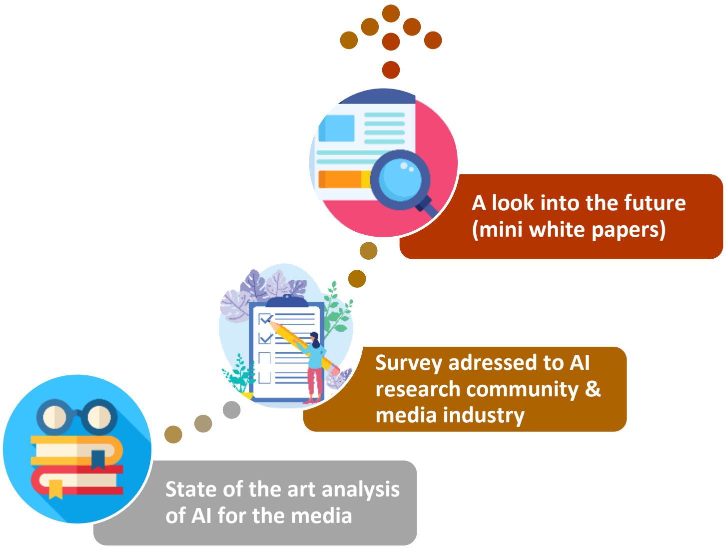


Figure 5: Tools used for the development of the AI4Media roadmap on AI for the media industry[33].

State of the art analysis: A large number of roadmaps, surveys, review papers, research papers and opinion articles focusing on the trends, benefits, challenges, and risks of the use of AI in different media industry sectors have been reviewed and analysed, aiming to provide a clear picture of the AI for media landscape. Our goal was to identify the most transformative applications of AI in the media and entertainment industry as well as the most important AI research trends and also provide insights on the main challenges (ethical, societal, economic, business, etc.) that the specific technologies and the adoption of AI by the media industry in general involve.

Survey: An online survey has been launched addressed to both AI researchers working on multimedia AI but also to people working in the media industry or whose work is closely related to this industry, aiming to collect their opinions on the benefits, risks, trends, challenges and ethics of AI use in the media. 150 responses were collected from AI researchers and media professionals from 26 countries in Europe and beyond and were analysed to identify main trends.

Look into the future: Based on the results of the state of the art analysis, we identified some of the most promising current and future trends in AI research and applications that could benefit the extended media industry. These technologies and applications were further analysed with a look into the future, aiming to provide information and insights on how they could positively disrupt the media and entertainment industry. For each identified technology/application, a mini white-paper was produced, providing information about the ***current status*** of the technology; ***research challenges*** relevant to its adoption and development; ***societal and media industry drivers*** for the adoption of the technology in the media sector; ***future trends*** presenting the potential applications of this technology in different sectors of the media industry; and ***goals***

[33] The icons used in this figure are from flaticon.com and vecteezy.com

for the next 10 or 20 years, summarising a set of milestones for the further development of these technologies and their application in the media. The same format is used for the development of similar white papers that focus on the deployment of AI on different media industry sectors, including news, social media, film/TV, games, music and publishing but also on the use of AI in the service of society, i.e. AI used to enhance online political debate, counteract disinformation or facilitate social science research.

One of the most interesting parts of this look into the future was the development of ***vignettes***, i.e. short stories with media practitioners or users of media services as the main characters, aiming to showcase how AI innovations could transform the media industry in practice.

This roadmap was the result of close collaboration between the members of the AI4Media consortium, with all research, technical and media industry partners contributing different subsections based on their expertise. In addition to these, this report also includes two contributions by invited external experts. More specifically, section 7.4 on AI fairness was authored by Samuel C. Hoffman from IBM Yorktown Heights in New York, an expert in algorithmic fairness and part of the team that develops the AI Fairness 360 open source toolkit[34]. In addition, section 9.9 on AI for publishing was authored by researchers participating in the Möbius H2020 project[35], which aims to modernise the European book publishing industry by remodelling the traditional value chains and business models, uncovering the prosumers potential and delivering new enriched media experiences. Figure 6 provides a visual overview of the structure of the road map and all the different topics we have focused on.

2.4 Structure of this report

The rest of this deliverable is structured as follows:

- **Section 3** presents the main findings and insights from a selection of roadmaps, surveys, review papers and opinion articles focusing on the trends, benefits, and challenges of the use of AI, aiming to provide a clear picture about the most transformative applications of AI in the media and entertainment industry but also of the most important current and future AI research trends that hold the potential for significant impact across the media industry value chain.
- **Section 4** summarises the results of the analysis of two surveys: a) a large-scale online survey aiming to collect the opinions of the AI research community and media industry professionals with regard to the benefits, risks, technological trends, challenges and ethics of AI use in the media industry ; and b) a small-scale internal survey addressed to AI4Media partners, aiming to collect their opinions on the benefits and risks of media AI for the society and democracy as well as on policies for the ethical use of media AI.
- **Section 5** analyses selected AI technologies, including white papers for reinforcement learning, evolutionary learning, learning with scarce data, transformers, causal AI, AI at

[34] AI Fairness 360: https://ai-fairness-360.org/
[35] Möbius project (funded by H2020 under grant agreement no 957185): https://mobius-project.eu/

the edge, bioinspired learning and quantum computing for AI learning, aiming to offer a clear overview of the current status of each technology, the drivers and challenges for its development and adoption, and future trends and goals.

- **Section 6** analyses selected AI applications for multimedia analysis, including white papers for multimodal knowledge representation and retrieval, media summarisation, automatic content creation, affective analysis, NLP applications and content moderation, aiming to offer a clear overview of the current status of each technology, the drivers and challenges for its development and adoption, and future trends and goals.
- **Section 7** focuses on four aspects of trustworthy AI (robustness, explainability, fairness and privacy), including relevant white papers that examine the current status, applications, challenges, and future trends for trustworthy AI for the media industry.
- **Section 8** focuses on AI datasets and benchmark competitions, including white papers that discuss current status, research challenges and future outlook, while also providing insights on the ethical and legal aspects that relate to this domain.
- **Section 9** focuses on the deployment of AI on different media industry sectors, including white papers discussing AI for news/journalism, social media, film/TV, games, music and publishing. It also explores the use of AI to address critical global phenomena such as disinformation, to enhance online political debate, and to help the study of media itself in the form of AI-enabled social science tools.
- **Section 10** discusses issues related to AI democratisation, including a white paper focusing on open repositories for AI algorithms and data as well as on the topic of integrated intelligence.
- **Section 11** examines ethical, societal, environmental and economic risks and concerns stemming from the adoption of AI in the media industry, including bias and discrimination, media (in)dependence, inequality in access, privacy, transparency, accountability, liability, labour displacement, misinformation as an institutional threat, and environmental impact. In addition, it provides a brief overview of existing EU policy and legal initiatives and their impact on future AI research for the media industry.
- **Section 12** discusses the forces that could shape the future of the use of AI in the media sector, focusing on legislation/regulation, the pandemic and its impact, and climate crisis.
- **Section 13** briefly presents the web version of this Roadmap, available at the AI4Media website.
- Finally, **section 14** summarises the conclusions.

This **Appendix (section 15)** presents the questionnaires used in the two AI4Media online surveys analysed in section 4.

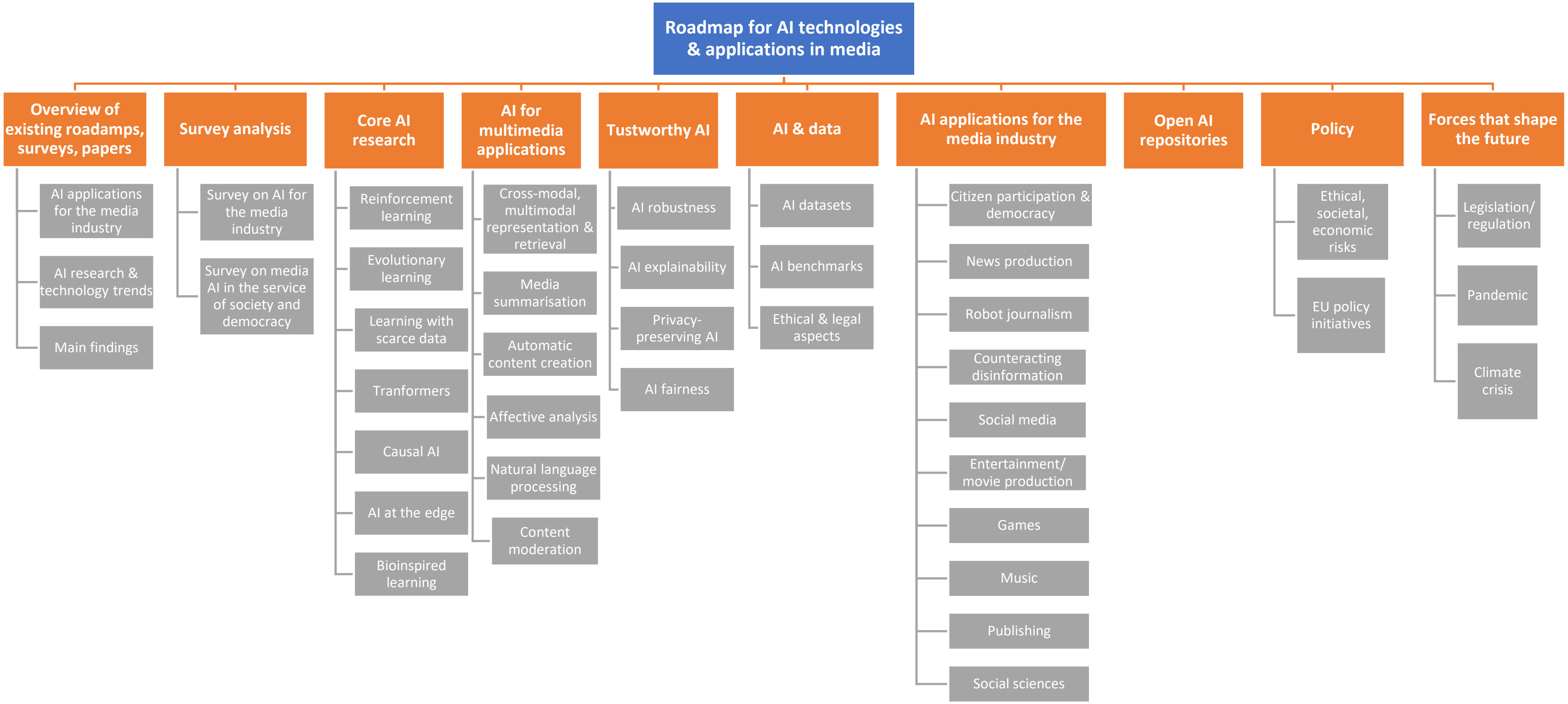


Figure 6: Overview of the structure of the roadmap on AI for the media industry and the various topics covered.

3 Overview of existing AI roadmaps, surveys and reviews

Contributors: *Filareti Tsalakanidou (CERTH)*

A large number of roadmaps, surveys, review papers and opinion articles focusing on the trends, benefits, and challenges of the use of AI in different industry sectors as well as in the public sector have been published during the last years. In this section, we briefly present the main findings and insights from a selection of these works, aiming to provide a clear picture about the most transformative applications of AI in the media and entertainment industry but also of the most important current and future AI research trends that hold the potential for significant impact across the media industry value chain.

This section is structured as follows: in subsection 3.1, we present the most important findings from selected surveys and studies that focus on AI application in the media sector. In subsection 3.2, we present the insights from reports that examine more generic AI technology and research trends that enhance and accelerate the science of AI in general. Finally, in subsection 3.3 we attempt to summarise the most prominent applications of AI for the media sector and also identify the most promising AI research trends that can be deployed to facilitate AI applications for the media sector.

This section does not claim to provide a comprehensive review of all literature on AI or AI for media, rather it is a selective review that focuses specifically on reports, papers and opinion articles that explore the potentials, impacts and challenges of AI in the media industry. The selected works were identified through a snowballing method where the involved AI4Media Consortium members identified core reports, which were considered good starting points for this overview. From these selected works, more reports, papers and articles were identified and considered based on their bibliographies. Furthermore, topic-specific searches were conducted for selected AI media applications or media sectors to ensure that some topics or areas are not underrepresented. For all these works, three criteria of relevance were determined: 1) they should explicitly deal with AI (either specific technologies or more widely), 2) they should either address directly AI for media or technologies typically used by media AI, and 3) they should offer insights regarding the impact, opportunities, future outlook and challenges of AI for the media industry.

3.1 AI applications for the media sector

In this section, we summarise the most important findings from selected surveys, reports and papers examining the use of AI in the media and entertainment industry. Our aim is to identify the most important applications of AI in the media sector, both existing but also future, and also provide some insights on the main challenges (ethical, societal, business) that these technologies and the adoption of AI in general involve.

3.1.1 AI.AT.Media – AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course (2021)

Contributors: *Werner Bailer (JR), Georg Thallinger (JR)*

A study titled "*AI.AT.Media – AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course*" was commissioned by the Austrian Federal Ministry for Climate Protection, Environment, Energy, Mobility, Innovation and Technology (BMK) in 2020. The aim was to identify the research potential of AI in the media sector in Austria, to describe thematic challenges, and to point out suitable options for actions for exploiting the AI potential. The final report (in German) was published in October 2021[36].

The methodology of the study consisted of a literature survey and web research, of a survey among media consumers (*n*=500), interviews with technology providers (*n*=19), media professionals (*n*=7) and researchers (*n*=9), and two half-day workshops bringing together participants from these three groups (mostly focusing on defining the challenges). In addition, a panel of three experts (IT, journalism, and law) provided inputs to the work.

AI application areas and automation levels

The study proposes a taxonomy of application areas, that extends and details of the one proposed in the "*World Press Trends 2020-21*" report[37]. As the top level structure, it uses the three main stages of the media value chain: ***sourcing*** deals with collecting, organising and assessing information; ***production*** is concerned with the actual creation of content, and ***distribution*** involves publishing, monitoring impact and interaction with consumers. The next level structures these stages into tasks or applications for certain types of content, with one additional level where needed. The granularity of the taxonomy was chosen based on the heterogeneity of the task and the amount of existing AI solutions (i.e., coarser groups were formed, where the available technology is still sparse). The study provides an assessment of the maturity of AI technology for each application area. The scale of this assessment goes beyond that of technology readiness level (TRL)[38], as it also assesses specifically the experimental or productive use in the media industry. The analysis shows a relatively high level of technical maturity in many applications (e.g., information extraction from text, visual classification, recommendation and content selection), with products and services being available, and experimental use being reported by media organisations. However, only a small share of applications (e.g., named entity recognition (NER) in text, automatic speech recognition (ASR) for large languages, content recommendation, and moderation of discussions) have moved into productive use beyond a few early adopters. This is mostly due to the lack of robustness, but

[36] V. Krawarik, K. Schell, V. Ertelthalner, G. Thallinger, and W. Bailer, "AI.AT.Media -- AI and the Austrian Media Sector: Mapping the Landscape, Setting a Course," Austrian Federal Ministry of Climate Action, Environment, Energy, Mobility, Innovation and Technology (2021): https://www.bmk.gv.at/themen/innovation/publikationen/ikt/ai/ai_at_media.html

[37] F. Nel and C. Milburn-Curtis: "World Press Trends 2020-21," Frankfurt: WAN-IFRA, the World Association of News Publishers (2021): https://wan-ifra.org/wp-content/uploads/2021/04/WAN-IFRA-Report_WPT2020-21.pdf

[38] Space systems - Definition of the Technology Readiness Levels (TRLs) and their criteria of assessment, ISO 16290, 2013.

also due to the lack of trained models on data relevant for the particular media organisation (e.g., non-English language models). Closing this gap is not feasible on the short term, and human-in-the-loop approaches are required to enable the gradual introduction of AI technologies in these application areas.

In the different application areas, AI may take a rather supporting role, being applied to very specific tasks, or have a higher degree of autonomy and contributes to decisions or makes them alone. The acceptable degree of automation depends very much on the application area, where media creators are usually more concerned about AI technology creating content or making decisions interfacing with customers. The study thus proposes five levels of automation (similar to the well-known automation levels for autonomous driving[39]: ***AI-enhanced tools***, ***AI-based assistance***, ***Conditional automation***, ***High automation*** and ***Full automation***. Descriptions of the levels and examples for different stages of the media value chain are provide in Table 1. For some applications in the media industry, the higher automation levels may be entirely out of scope, or limited to specific cases. One reason is the technical feasibility, as journalism's task is to report about ``the world'', i.e., news may involve all possible domains and topics. Thus, the requirements for a fully automatic solution may come close to that of artificial general intelligence (AGI). Another important reason is the wish to keep human oversight over information and processes that may have a strong impact on democracy and society, rather than leaving automated content generation and recommendation algorithms to negotiate public opinion.

Table 1: Automation levels and examples for different stages of the media value chain.

Level	Description	Sourcing	Production	Distribution
Level 0 - No automation	Humans control non-AI-based tools	Topics monitored by humans, content analyzed and verified by humans	Content created by humans, tools based on non-AI technologies	Content selection and playout by humans
Level 1 - AI-enhanced tools	Humans operate tools using AI for specific tasks in their workflows	Content quality analysis, content similarity/near duplicates, model fitting, prediction	Content modification and enhancement tools use AI for low level tasks (e.g., colour correction, spell checking, inpainting)	Media monitoring tools
Level 2 - AI-based assistance	AI-based tools generate information that is used in subsequent steps, with human verification/correction, no decision	ASR, content tagging, classification, object/logo/face detection, trustworthiness scoring,	Content suggestion/completion, summarisation, subtitling support	Suggestion of content and ads

[39] Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles. SAE J3016_202104, 2021: https://www.sae.org/standards/content/j3016_202104/

Level	Description	Sourcing	Production	Distribution
		knowledge modelling		
Level 3 - Conditional automation	Processes including AI-based decisions, with human intervention required at some points	Selection of relevant topics and sets of source content	Text/media generation from highly structured information, preparation of accessibility content	Automated choice of encoder settings, bitrate selection, content recommendation, automated compliance checking
Level 4 - High automation	Fully AI-based processes under human supervision (in particular for consumer facing decisions)	Automatic filtering and selection of sources, automatic content verification	Automatic generation of content (with review), fully automatic generation of versions	Automated user targeting and content adaptation to user (with human checks), chatbot with handoff
Level 5 – Full automation	Fully AI-based processes directly interfacing consumer, without human supervision	Automated relevance assessment, and analysis, assessment of content	Automated content creation/adaption tools for all modalities, automatic accessibility	Automated user targeting and content adaptation to user, chatbot without handoff

Key challenges identified

In terms of relevant emerging research fields, the study highlights ***question answering and captioning***, ***media verification and forensics***, ***content generation and improvement***, ***learning-based media coding***, ***moderation and discourse analysis***, ***AI explainability and transparency*** as well as ***learning with scarce data***.

The study identifies four key challenges for making AI more widely usable in the media industry. First, the availability of ***localised AI tools and data***. Many solutions work well for English, but less so for smaller languages, and not at all for local dialects. Apart from this issue with language processing resources, a similar problem exists with the ***availability of (linked) open data for regional or local matters***, as well as *image content* (e.g., for training landmarks or face recognition). Second, ***content generation tools*** have not yet matured beyond filling templates. Any other media generation does not yet meet the quality expectations of media companies and audiences. In addition, content generation tools need to meet journalistic standards in order to be usable, i.e., ensure that generated content is factually correct, that information and opinions about it are clearly separated, etc. The third challenge concerns ***better targeting,*** *respecting the consumers' privacy and avoiding filter bubbles*. Especially smaller media companies lack data on their consumers, and drilling further down in small audiences raises privacy issues. Fourth, ***assistive tools*** specifically targeted to journalistic needs are required, providing support in standard tasks in content sourcing and creation (e.g., summarising the state of the story so far, before giving further updates). While some of the technology for these tasks is already available, it often does not meet requirements of journalistic workflows, e.g., ensuring that the source of every piece of information provided can be traced back and easily checked.

3.1.2 The AI Disruption in Media & Entertainment (Dataiku, 2020)

This survey[40] by Dataiku offers a comprehensive overview of the ways in which AI can disrupt and transform the media and entertainment industry. This is achieved through the presentation of high value AI use cases for the media, key AI opportunities per media sector, and main challenges and upcoming trends for AI in media. The central themes running through the report are the need for the adoption of data-driven business models and the trend of personalisation of services that allow media companies to expand their business by offering their content "not at audiences of billions, but at billions of individuals".

The report offers a comprehensive list of **high-value AI-driven use cases** for the media industry, which includes AI applications like:

- ***Smart recommendation engines***: AI can enable the shift from simple recommendation engines into fully personalised content experiences, improving content relevance and distribution.
- ***Hyper-targeted advertising***: By combining and analysing large volumes of data, AI can enable accurate prediction of user churn rates, steer advertising placement, and increase conversion through personalised offers like personalised movie trailers appealing to different audiences.
- ***Real-time predictive modelling for anticipating demand and segmentation***: This can help media companies react in real time to consumers' changing needs, predict future needs to direct investments accordingly, and anticipate content engagement per audience segment.
- ***Programmatic ad buying***: Based on real-time analysis of audience dynamics across multiple channels and ad space, the process of ad buying on different media platforms can be efficiently automated.
- ***Personalised programming***: By collecting and analysing a steady stream of customer data, streaming companies can offer not only personalised program recommendations based on sophisticated user profiles but also personalised formats for the presentation of the content.
- ***Automated "robot reporting"***: NLP technologies are already used to automatically produce news content for topics like financial news or minor-league sports but also to detect trends in social media thus identifying topics that may interest the audience. AI can be exploited to assist (or replace) journalists in tedious or boring tasks, allowing them to focus on more meaningful work like news analysis or investigative reporting.

[40] The AI Disruption in Media & Entertainment (2020): https://content.dataiku.com/ai-media-entertainment/ai-media-entertainment

- ***Video game development***: New AI technologies like evolution learning are already transforming the industry, helping design more sophisticated games with virtual characters that can think and act like humans (reacting to or anticipating the human players' actions).
- ***Automating HR, legal and administrative tasks***: Apart from AI functionalities to improve offered services and expand business, media companies can also exploit AI advances to automate administrative, legal or HR tasks. Some prominent examples include *forecast of residual payments* to anticipate or explore scenarios of how talent (e.g. actors in a movie, music creators, etc.) will be compensated based on content distribution across platforms; *compliance with copyright standards* by automatically analysing relevant legal documents; *AI-based casting* for TV shows/films based on cross-analysis of historical data of film performance and relevant cast profiles and prediction modelling.

In addition to identifying high-value use cases along the media production chain, the report also highlights **how AI can help individual media & entertainment sectors**:

- **Music**: AI has already disrupted the music industry by offering innovative solutions in areas such as recommendation engines for content personalisation, forecasting of music trends & sales, segmentation of customer base based on audience and user behaviour analysis, forecasting for content monetisation and payments to talent.
- **Film & TV**: AI offers the potential of radical transformation especially for film studios, which unlike streaming services have not fully embraced yet the data-driven model. AI-enabled analysis of user behaviour and operational data can revolutionise content recommendation systems, churn prediction, hyper-targeted marketing, sale and residual payment forecasting, but also analysis of sentiment and trends is social media for market research.
- **Video games**: AI can facilitate game development, design and graphics, aiming to produce sophisticated games and maximise user engagement while it can also support personalised marketing of games and social and customer analysis, aiming to maximise revenue, understand the players, and maximise their satisfaction.
- **News**: AI can help to reinvent the news industry model through applications like robot journalism, hyper-targeted advertising, sales forecasting etc. Interestingly enough, the report does not delve into more journalistic aspects examined in similar works (e.g. AI for countering disinformation, for automated news production or coverage, or for content archiving and search).
- **Sports entertainment & gambling**: Aiming to improve the online gambling experience, AI is used not only for personalisation or user behaviour prediction but also for modelling the uncertainty surrounding user actions. Prominent applications of AI for this sector include fraud detection, lifetime customer value prediction, and ad-hoc analysis of AB/test data.
- **Advertising**: AI is already transforming the way ads are bought or sold by enabling hyper-targeted advertising, programmatic ad buying, and customer base segmentation.

The use of AI in the media sector also comes with a set of **challenges**, including:

- Creating ***sustainable and effective production pipelines*** to process and exploit data.
- Issues of ***data governance, data privacy and compliance with relevant regulations***. The need for advanced collection of personal data (even sensitive) to support personalisation of media experiences often clashes with relevant regulation.
- ***Misinformation in social media***. The business model of social media, which is based on maximising user engagement/reaction, creates well-founded worries in the news industry since it has been shown to significantly contribute to misinformation spread.
- ***Recommendation engines and filter bubbles***. While recommenders are fundamental in helping to maximise user engagement by providing users content they like, they are also responsible for creating filter bubbles by feeding users news content biased towards their beliefs, thus failing to provide the bigger picture or alternative views and isolating users in an eco-chamber. AI can be used to improve recommender systems by increasing transparency and identifying biases.
- ***Attracting AI talent***. Media companies have a different set of unsolved problems and thus different needs when it comes to hiring AI talent compared to big tech companies.

The report concludes by nicely summarising the **main trends in AI and ML for the media sector**:

- ***More personalisation***. Media experiences will become highly personalised. In addition to content personalisation, the trend is towards integration across media experiences and also use of AI assistants to facilitate content discovery.
- ***Convergence of technology, telecom and media industry***. Tech and telecom companies acquire media companies in a fast pace, driven by increased demands for data and AI tech.
- ***ML to enhance user experience***. Based on use of text and image analysis, advanced recommender systems will suggest images for news articles or search for images of the same type. At the same time, AI will facilitate multi-lingual content translation as well as transformation to different writing styles for different audiences.
- ***AI for immersive VR experiences***. AI will facilitate the creation of fully immersive AR/VR experiences based on intelligent avatars and VR content created by AI.
- ***AI beyond content***. Media companies have already started adopting AI technologies for administrative, legal, HR or other tasks.

3.1.3 The Technology, Media & Telecommunications AI Dossier (Deloitte, 2021)

This report[41] by the Deloitte AI Institute highlights the most business-ready cases of AI for the Technology, Media & Telecommunications (TMT) sector while also presenting emerging AI

[41] Deloitte AI Institute, The Technology, Media & Telecommunications AI Dossier (2021): https://www2.deloitte.com/content/dam/Deloitte/us/Documents/deloitte-analytics/us-ai-institute-dossier-tech-media-telecomm-dossier.pdf

applications with future potential. Below, we briefly discuss those most relevant to the media sector. Current applications include:

- ***Increased user engagement***, using AI to automate engagement and two-way communication with users through NLP and sentiment analysis but also to improve content and services personalisation via AI-powered DM platforms.
- ***Digital Contact Centers,*** using NLP to build virtual assistants that deliver a more human-like communication but also adopting predictive analytics and sentiment analysis to monitor interactions and provide useful insights for customers and staff.
- ***Fake media content detection***, based on video and text analysis to detect deepfake videos but also other media content like fake articles.
- ***Monetisation of customer data,*** based on analysis of user behaviour (e.g. conversations on social media) to provide content they like or advertise products they are likely to buy but also based on the combination of different sources of user data for more efficient targeting.

Future trends include:

- ***Language translation services***, using NLP technology for automatic real-time translation of content and elevated communication experience, free of language barriers.
- ***Video content analysis.*** Use of ML, DL and computer vision to automatically analyse video content and thus facilitate tasks such as real-time monitoring and trend detection in social media aiming at increased personalisation but also enable monetisation of video archive collections.
- ***Audio and video mining,*** using AI to transform video and audio in structured data and thus easily mine the vast volumes of user generated content for useful information.
- ***Ad analytics based on emotion detection:*** using data collected by sensors installed in the viewer's living room, the film/TV industry as well as advertisers will be able to understand how the viewer is emotionally affected by the content presented to them and will thus be able to extract useful insights about user preferences or predict future engagement with content.
- ***Self-healing networks***, using AI-enabled predictive analytics to predict network maintenance needs, e.g. for broadcast infrastructure, thus minimising failures and costs.

3.1.4 Artificial intelligence systems for programme production and exchange (ITU-R, 2019)

This report[42] by the Radiocommunication Sector of the International Telecommunication Union (ITU-R) of the United Nations presents an overview of **AI applications across the whole value chain of the broadcasting sector**, from workflow optimisation to personalisation of content:

[42] Artificial intelligence systems for programme production and exchange, Report ITU-R BT.2447-0 (2019): https://www.itu.int/dms_pub/itu-r/opb/rep/R-REP-BT.2447-2019-PDF-E.pdf

- ***AI for workflow optimisation,*** aiming to automate tedious routine tasks. This includes analysis of archived content and historical scheduling data for optimising content programming for different audiences; automatic camera selection and framing; automated content generation, e.g. creation of highlights for sports events; automated creation of video digests based on image analysis of content and analysis of user comments on social media; optimisation of live footage based on learned relationship dynamics of group interactions, etc.
- ***Automated content creation***, including news alerts based on analysis of social media posts (text/image/video) to detect newsworthy trends or trustworthy news; automatic generation of content based on analysis of data released by different agencies or organisations (e.g. financial data, weather data, etc.); automatic captioning based on speech recognition; chatbots to dialogue with users/audience; automated AI-driven announcer systems; automated commentary for live programmes like sports events; translation in different languages; sign-language CG synthesis; automated programme creation based on analysis and synthesis of archived data.
- ***Metadata creation*** to optimise content management and search in vast audio-visual broadcast archives. For the creation of high-level metadata, AI algorithms for video and audio analysis and detection, face detection and recognition, detection of text in video (e.g. detection of signs), object detection and recognition, and speech recognition are used.
- ***Dynamic product placement and advertising***. Use of AI technologies that will decouple product placement from the initial content generation. E.g. placement of different products (e.g. a drink from a different brand) on a broadcast film depending on audience or timeslot will be able during post-production and distribution.
- ***Content personalisation***. Personalisation of content has become significantly important in broadcast to target efficiently audiences of different demographics. More recent trends in this direction include user-decision-driven modular storylines as well as personalisation of content based on the user's affective state, captured by real-time sensors.

The report concludes by highlighting the need for large volumes of real and high-quality data for training AI models for the media industry.

3.1.5 AI in the media and creative industries (NEM, 2019)

This white paper[43] by the NEM initiative aims to offer a detailed overview of AI technological advances and their potential impact on the media sector, also identifying relevant challenges. The paper is structured upon five creative application areas (**music, images & visual art, digital storytelling**, etc.) and three axes (**creation, production and consumption of media content**). In

[43] AI in the media and creative industries, New European Media Initiative (2019): https://arxiv.org/ftp/arxiv/papers/1905/1905.04175.pdf

Table 2 below, we attempt to briefly summarise the findings for each application area. As can be seen, the authors of this report go into significant depth on the ways in which AI can be used to achieve different tasks in the selected creative areas. The report offers also details on the specific AI technologies used in each case (e.g. GANs, CNNs, etc.).

Table 2: Overview of AI technologies for the media industries and relevant challenges in NEM white paper on "AI in the media and creative industries"

Application areas	Content creation	Content production	Content consumption	Challenges
Music	- Generative models able to handle advanced audio features such as timbre - Style transfer - Tools for personalised music generation	- AI-based music production systems for independent musicians - Audio demixing - Processing of different music signals at the same time without losing audio quality	- Personalised recommendations - Improved playback allowing user control over sound - Delivery of personalised music content	- Models that leverage information in raw audio signals -Learning long-term temporal dependencies - Transfer between different timbres - Denoising and demixing
Images and visual art	-Apps for AI generated art images - Generation of pastiches - Generation of images from captions	- Reproduction of scene-dependent image transformations - Image quality enhancement - Style-based image transformation - Image inpainting - Image retrieval based on visual features or text	- Automatic metadata creation based on image/text analysis - Object/face detection &recognition	- Image generation from description - Understanding image content - Cross-media search - Automatic reconstruction of missing parts
Digital storytelling	- Searching archives to find suitable content to fit the narrative -Sensor-based storytelling for immersive experiences - AI cinematography - Authoring tools for non-experts - Transmedia storytelling		- Multimodal interactive and virtual experiences - Content personalisation - Audience segmentation	-Transdisciplinary storytelling bringing together writing, tech and gaming - Automation of narration in interactive entertainment - Ensuring authors remain in control of stories partly created by AI

Application areas	Content creation	Content production	Content consumption	Challenges
Content in games, movies, engineering and design	- Automatic content synthesis (detailed landscapes, textures, objects, etc.) - Generative game design - By-example synthesis		-Personalised games adapted dynamically to player profiles - Modelling player behaviour, skills and affective state	- Synthesis of large environments - Methodologies that directly sample instances by directly enforcing requirements - Reliable estimation of user emotional state -Ethics by design to prevent addiction
AI and Book Publishing	- Automation of processes in publishing houses - Improved accessibility for people with print impairments (e.g. automatic text creation to describe image concept) - Publication tagging - Copyright management (e.g. using visual and text info from the web to provide relevant licensing info, combination of AI and blockchain) - Book recommendations based on book content analysis and user profiles			-Addressing IPR issues related to AI input and output
Media access services	- Automatic subtitling (contextualised, multi-lingual, complying with legislation) - Sign language production - Audio description of content			- Streamline circulation of audio-visual content through translation - Automatic multilingual translation
News	-Analysis and cross-examination of various information sources in multiple languages - Fact-checking - Content verification (detection of manipulated content/ deepfakes) - Social network analysis to monitor disinformation spread - Automatic creation of personalised news digests		- User profiling & recommender systems - Public tools for assessing information reliability - News aggregation and summarisation -Intelligent tools for high-quality participative journalism	- Heterogeneous data integration and search - Security of multimodal information retrieval systems and video forensics - Trust and transparency of AI algorithms
Social media	- Content personalisation - Recommender systems - Content search	- Ad placement - Trend detection - Moderation		

Application areas	Content creation	Content production	Content consumption	Challenges
	- Opinion mining			

Apart from the challenges identified for each media/creative sector, the NEM white paper also identifies a set of transversal challenges across sectors, summarised below:

- *Data challenges*, including building large annotated databases not only of content but also of user preferences.
- *Robustness challenges* with regard to performance of pattern recognition algorithms.
- *Cross-domain challenges*, i.e. development of cross-domain multi-modal analysis tools.
- *Challenges of human-machine collaboration,* including AI explainability and adaptation of AI tools to user needs.
- *AI transparency and accountability challenges*.
- *Ethics challenges*, including development of ethics-by-design tools as well as improved AI transparency and trustworthiness.

3.1.6 Artificial Intelligence in the Creative Industries: A Review (BVI, 2021)

This paper[44] by the Bristol Vision Institute of the University of Bristol offers an overview of the state-of-the-art for the use of AI technologies in the creative industry. The authors classify AI applications in five areas (***content creation, information analysis, content enhancement & post production workflows, information extraction & enhancement,*** *and* ***data compression***). For each area, several topics are identified and examples of AI applications are provided for each topic. Unlike most of the works presented in this section, this survey focuses a lot on the actual AI technologies deployed for each application, i.e. CNNs, GANs, RL, BERT, VAE, etc. In the following Table, we summarise the different examples of AI application in the creative industries classified per area and topic.

Table 3: Summary of AI application areas in the creative industries from the "Artificial Intelligence in the Creative Industries: A review", a survey by BVI.

Application area topic	Examples of application
Content creation	
Script and movie generation	• Script generation for film • Automated generation of movie trailers • Interactive narrative for games • Procedural content generation for games
Journalism and text generation	• Robot journalism for generic news • Language translation • Video to text • Text transformation to different writing styles
Music generation	• Search in audio databases • AI-assisted music creation
New image generation	• Automatic image creation

[44] N. Anantrasirichai and D. Bull, Artificial Intelligence in the Creative Industries: A Review (2021): https://www.researchgate.net/publication/343228503_Artificial_Intelligence_in_the_Creative_Industries_A_Review

Application area topic	Examples of application
	• Image transformation/style transfer (e.g. from one painter style to another, face aging, etc.) • Talking video from facial image
Animation	• Character animation for film and games • Scene and object rendering for film and games • Human-like virtual assistants
AR/VR/MR	• Game design • User motion detection • Dynamic virtual environments for film and games • Object and user detection
Deepfakes	• Improved realism in film industry (e.g. replacement of one actor's face with another) • Manipulated/fake audio/video/text for disinformation • Manipulated/fake digital content detection
Content and captions	• Text generation from image/video analysis/interpretation • Image generation from text analysis
Information analysis	
Text categorisation	• Document indexing and retrieval • Sentiment classification • Topic classification • Spam detection
Advertisements & film analysis	• Recommender systems for music and movies • User profiling/ behaviour analysis • Dynamic ad programming/placement • Social media analysis for opinion mining • Content performance prediction (e.g. by historical data analysis that associates box-office performance with film content)
Content retrieval	• Automatic annotation and metadata creation (based on audio/object recognition and scene understanding) • Image/video search based on image analysis and understanding • Music retrieval based on search by sound, query by humming, etc.
Recommendation engines	• Content-based filtering based on single user preferences • Collaborative filtering based on other user suggestions • Knowledge- based filtering based on user queries • Content summarisation
Intelligent assistants (based on text or voice)	• Information retrieval (e.g. relevant news, weather reports, etc.) • Recommendation of content • Chatbots for dialogue with audience
Content Enhancement and Post Production Workflows	
Enhancement	• Contrast enhancement • Colourisation to add or restore colour in images • Super-resolution images
Post-production	• Image deblurring, denoising, dehazing/mitigating atmospheric turbulence • Inpainting (e.g. for removal of unwanted objects or restoring missing parts) • Visual Special Effects to create realistic 3D virtual characters, animations, environments

Application area topic	Examples of application
Information Extraction and Enhancement	
Segmentation	• Object detection & classification • Semantic segmentation (in 2D and 3D)
Recognition	• Object recognition • Face recognition • Speech and music recognition • Emotion detection (based on video, audio) • Action recognition • Sign language recognition
Tracking	• Object and face tracking in video
Image fusion	• Combination of multiple images to aid human perception
3D reconstruction & rendering	• 3D scene reconstruction from video or RGBD sensor output
Data compression	
Data compression	• Deep compression methods to optimise existing coding tools • Compression for super-resolution videos

The review also points out ethical issues related with the use of AI in the creative industries such as ***authorship of AI creations***, ***AI-induced inequality*** (with regard to information access*), **use of AI with malicious intent*** (e.g. for creation of deepfakes aiming to spread disinformation), and ***AI algorithmic bias***. It also stresses that having a **human in the loop** is necessary for current AI systems, especially for creative processes. This necessitates the establishment of a feedback mechanism that will allow humans to check AI outputs, make critical decisions based on AI output, and provide feedback about AI failures to the AI system. While the most effective AI algorithms are still based on ground truth and labelling, in truly creative processes there is usually no way to evaluate the quality of the generated outcome in advance. Thus new methods going beyond traditional ML algorithms, e.g. methods like generative models, should be further explored and extended to generate new creative content.

With regard to the future of AI, the report singles out the following areas that hold great promise but require further research efforts:

- Increased diversity and context in AI training for new ***creative content generation***. This will require high-dimensional datasets including info on audience preferences and current trends but also modelling human perception of quality;
- ***Convergence of AI with blockchain*** technologies to support improved and trusted data labelling and model training;
- Techniques for ***unsupervised or self-supervised learning*** will be significantly important as the volume of collected data grows exponentially;
- ***Reinforcement and transfer learning*** to provide greater generalisation capabilities for AI learning algorithms;
- Research on ***human learning mechanisms***, aiming to imitate human brain capabilities.

3.1.7 New powers, new responsibilities - A global survey of journalism and artificial intelligence (LSE, POLIS, Google News Initiative, 2019)

This report[45] by London School of Economics, POLIS and Google News Initiative is based on a survey of 116 journalists in 71 news organisations in 32 countries focusing on AI for journalism. The report finds that AI is already used in newsrooms for a variety of tasks, mainly for automating tedious tasks and supporting journalists in creating better and more trusted journalism; but also for delivering more relevant and useful content and services to users, helping citizens to cope successfully with information overload and misinformation rise. The main three areas where AI can improve newsroom functionality include ***automated tagging/entity extraction*** (for newsgathering), ***automatically generated content*** (for news production), and **enhanced recommendation engines and content personalisation** (for news distribution).

At the same time the respondents identified several challenges related to the use of AI in the newsroom: ***algorithmic bias*** leading to bad editorial decisions or discrimination against groups of people; ***disinformation spread and filter bubbles*** - AI is instrumental in creating these phenomena but it can also help newsrooms counter them; AI can help increase ***transparency*** as well as increase diversity of stories and audiences; ensuring ***balance between AI and human intelligence***; and the power of Big Tech companies that control development of AI technology. Overall, the newsrooms were confident that the impact of AI could be beneficial, given that they would retain ethical and editorial principles.

Half of newsrooms saw themselves as AI-ready while the other half were just starting to use such technologies and were already fearing that they were falling behind (especially small newsrooms). The main challenges for adopting AI include financial resources to build/manage AI systems, need for personnel with AI skills, cultural resistance (e.g. fear for loss of jobs or hostility against technology) but also lack of AI literature or clear AI strategy in the company.

Interestingly, the role of AI is characterised by the respondents as supplementary or catalytic but not transformational (yet). With regard to the future, the report identifies nine ways in which AI can reshape newsrooms and journalism:

- Improved content personalisation;
- Automated production of content;
- Dynamic pricing for ads and subscriptions;
- Automated transcriptions;
- Improved content moderation;
- Detection of fake news and deepfakes;
- Debunking of information;
- Enhanced content search based on video/image analysis;
- Sentiment analysis of user generated content.

In this direction, some interesting ideas collected from the journalists participating in the survey include: *automatic text to anything* (voice, video, other text); production of news enriched with more sources and personal aspects; *journalism of things*, exploiting everyday devices to collect

[45] C. Beckett, New powers, new responsibilities - A global survey of journalism and artificial intelligence (2019): https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities/

data (e.g. sensors in a political event to record crowd reaction to parts of the speech, or in sports to detect the most important highlights of a game); *augmented journalism* using drones, wearables, voice, VR technologies for novel content creation and delivery; transforming news from unidirectional communication to bidirectional communication through *enhanced user understanding* (incl. emotion analysis) and increased user engagement; working with news patterns rather than case-by-case stories; *conversational agents* that can reliably answer questions about current affairs.

The need for improved AI training and education was an issue of unanimous consent. Respondents pointed out the need for: **improved AI literacy** (what is AI and how it works) across a news organisation; **basic AI training** (e.g. basic skills on AI coding or data model training) for newsroom employees; innovation training through **experimental AI projects**; raising **awareness about AI** use in the industry and what other newsrooms do; **understanding AI ethical concerns** (bias, robustness, etc.) – how they work and how to address them; enhanced understanding of technology big picture and of wider role/impact of AI on society.

Although the news industry is highly competitive, the appetite for collaboration with regard to AI technology is notable. This collaboration may take various forms: across different units of the same organisation; between news organisations; between news industry and tech companies, start-ups or research/academic organisations; on a national or international level (e.g. to build transnational tools, adaptable to different audiences, languages, etc., or stories, e.g. on climate change, crime, etc.). Such collaboration can have a positive impact not only on news content and stories, but also on the development of novel AI solutions tailored for newsrooms and on financial costs.

The report concludes with how the news industry can learn from other industries when it comes to the adoption of AI: the retail industry can offer useful examples of recommendation engines, dynamic pricing or customer experience analysis; the gambling industry can offer examples of user behaviour analysis and audience understanding; the gaming industry offers examples of how to use AI to automatically interact with the users; medicine and biotech companies offer examples of standards for AI ethics; and of course big tech companies can offer insights on the next big trends of AI but also examples of success stories and failures/mistakes with regard to market and ethical challenges.

3.1.8 AI predictions 2021 - Technology, Media & Telecommunications (pwc, 2021)

This pwc survey[46] offers a good overview of the top priorities, benefits and challenges for the deployment of AI technologies by the tech, media and telecom industries. The survey unveils that the media companies' top goal with regard to the use of AI is to increase efficiency and productivity (53%), followed by revenue growth (45%), innovative products/services (34%), improved internal decision-making (29%), employee training (29%), and better user experiences (26%).

[46] AI predictions 2021 - Technology, Media & Telecommunications: https://www.pwc.com/us/en/tech-effect/ai-analytics/ai-predictions/technology-media-and-telecommunications.html

With regard to challenges, media companies identify as top challenge the development of AI models and datasets to be used across the company (47%), followed by recruitment of skilled personnel to work in AI (32%). Training of current employees (24%), making AI responsible and trustworthy (24%) and measuring AI's return on investment (21%) are deemed as important only by one fourth of respondents. At the same time, 45% of media companies report that they fully address risks related to AI and add necessary controls while 47% have a plan in place to identify new roles required as a result of the adoption of AI technology.

The survey concludes with three main takeaways about what should media companies do with regard to the adoption of AI:

- Take advantage of all the emerging AI technologies to **transform their workflows**, **improve consumer experience**, **forecast demand for content and services**, and **improve their marketing**.
- Operationalise AI across the organisation by creating a new operating model, adopting a new approach to technology, and new ways of work.
- Minimise AI-related risks, by updating **data governance, data policy and security** as well as by addressing issues of **AI bias** in their models.

3.1.9 Selection of online articles on AI applications and trends for the media sector

In the previous subsections, we briefly presented the findings of roadmaps, studies and surveys that aim to offer an accurate glimpse into the future of AI and it application in the media sector. These works usually follow a structured approach involving a co-creation process that aims to collect the opinions of different AI experts (or the public in some cases) and then synthesise these opinions to provide some insights and proposals about the future of AI research and applications. In addition to these works, in this subsection we briefly present the main AI trends and challenges identified in a selection of interesting articles posted by individual AI experts or groups of experts and published in relevant popular online magazines, websites, blogs, etc.

3.1.9.1 How Artificial Intelligence is transforming the media industry? (CMF Trends and Méta-Media, 2019)

This two part overview by CMF Trends and Méta-Media[47,48]presents a wide list of AI applications covering the whole information and entertainment media value chain. Some examples are presented below:

- ***AI for augmented information***: 'Robot' journalists can be used to speed up the news production process, however their use is currently limited to financial reporting, election reporting or small-scale events. Advances in NLP can help extend robot journalism to cover more complex topics, by identifying relevant content, context and appropriate presentation format. In addition, AI can help journalists to analyse vast amounts of data by combining

[47] K. Bremme, How Artificial Intelligence is transforming the media industry? – Part 1: https://cmf-fmc.ca/now-next/articles/how-artificial-intelligence-is-transforming-the-media-industry-part-1/
[48] K. Bremme, How Artificial Intelligence is transforming the media industry? – Part 2: https://cmf-fmc.ca/now-next/articles/how-artificial-intelligence-is-transforming-the-media-industry-part-2/

multiple sources of information. It also enables monitoring trends in social media and providing relevant automated news updates. To remain relevant, smart collaboration is required between journalists and AI assistants.

- ***AI for countering fake news***: AI plays a significant role both in the generation and spread of disinformation but also in the detection of disinformation. AI can be used to automate fact-checking of information and verification of content using advanced text, image, video and audio analysis.
- ***AI for improving online conversations***: NLP technologies are used to moderate online conversations, by identifying hate speech, verbal violence, etc. In addition image analysis is used to detect hate-filled or violent images. Such automatic moderation techniques are especially helpful in cases where the lack of human moderators and the toxic comments have obliged editors to turn off comments on sites and online articles. Using automatic moderation tools will give the audience the opportunity to express themselves in a civilised online environment.
- ***AI at the service of voice***: NLP and voice recognition have led to conversational agents capable of understanding humans and dialoguing with them. AI enables high quality text-to-speech synthesis, speech-to-text translation and support for different languages and dialects. These capabilities are extremely helpful for the whole media industry.
- ***AI for interactivity and engagement***: Many online news sites use such bots to interact with their audience and share news, while interactive fiction content has also become a thing, e.g. with interactive stories for children or adults. The conversational and personalisation capabilities of chatbots create a closer user experience, enhance interaction with the user and foster engagement. This type of AI technology can provide new storytelling experiences and increase user engagement in sectors like advertising, marketing, film and audio.
- ***AI in XR***: AI has the potential to advance storytelling with virtual characters that are capable of advanced interactions with human beings. AI can augment virtual character design (facial expressions, body movements, voice), can identify human emotions and can enhance interaction between humans and virtual characters by making it more natural.
- ***AI for indexing, archiving and search***: by combining image analysis, NLP and ML, AI can automate metadata creation for multimedia content thus enhancing archiving and discoverability. AI can also automate other content management tasks like data format conversion or sub-title extraction, thus enabling real-time indexing. Automatic metadata creation more effective content search, increases monetisation opportunities, helps media practitioners in their daily tasks (e.g. reporters to search video archive content to write a news story or fact-check a piece of information).
- ***AI for targeting and customisation***: AI can revolutionise recommendation algorithms and help provide the right content to the right person at the right time, based on user profile and activity, also considering the context (e.g. place, time, weather). Most of media already use AI to recommend content while giants like Amazon or Netflix heavily rely on personalisation of services.
- ***AI for accessibility***: AI can help make content more accessible for people with disabilities or limitations, e.g. by automating subtitles production, text-to-speech technologies, image recognition for audio description, and real-time translation.

- ***AI for video production and creation***: AI can be used for automated video editing and creation. For example, lengthy videos can be analysed to create short versions with highlights (e.g. short summaries tailored for social media use) or produce automatically edited versions ready for distribution. AI can also be used for post-production to offer different montages of a scene or add special effects.
- ***AI for monetisation and prediction of success***: based on techniques such as analysis of behavioural data, audience analysis and trend detection, AI can predict the success or engagement of content before this is made available to users. To this end, AI is present in the whole marketing chain: from customer acquisition (audience analysis and segmentation, scoring and targeting, visual context identification), to transformation (customisation and recommendation, content creation, optimisation of sites and media, automated campaign piloting), and loyalty building (conversational agents, customer program automation, behaviour analysis, predictions). Another more recent trend involves the use of AI-enabled emotion recognition that will also allow content recommendation based on the user's emotional state at any given time.
- ***AI and media ethics***: issues such AI bias, data and user privacy, AI interpretability and explainability are challenges that should be addressed to ensure ethical use of AI in the media sector.

The article concludes by pointing out the potential of AI for content access and monetisation, personalised recommendations and manipulation prevention, but also stressing the need for collaboration of humans and AI when it comes to content creation or management.

3.1.9.2 Transforming the media industry with AI (The Record, 2018)

This article[49] featured in The Record magazine tries to explore how AI can transform the entire media and entertainment industry from content creation to user experience. Based on interviews with experts that either provide AI technologies for the media sector or analyse the media sector, the following AI technologies with a potential to bring transformational changes in different parts of the media business chain are identified:

- ***Automatic metadata tagging and extraction*** based on image analysis and speech-to-text technology. This feature can help media companies effectively organise their content and archives and is going to be extremely useful for driving content exploitation and relative monetisation strategies.
- ***Use of AI tools to strengthen predictive capabilities***, e.g. to predict content demand and subsequently adjust resources in the cloud or predict disruption to supply chains, thus ensuring considerable savings for companies.
- ***Personalised content distribution*** to media consumers, including title recommendations or content curation based on user profiles which are built by collecting data about user interactions with content in the cloud.

[49] L. James, Transforming the media industry with AI: http://digital.tudor-rose.co.uk/therecord/issue09/64/

- ***Personalised media advertisement campaigns*** based on AI that autonomously understands product properties, consumer preferences and willingness to pay per product, purchase of likelihood at a given time, etc.
- ***Automation and digitalisation of existing workflows***, enabling content production and distribution to meet growing demands.
- ***AI-enabled feedback for artists*** based on real-time metrics of consumption that has the potential to increase creativity and increase production of content.

The main challenge identified for the realisation of AI's potential for the media sector is the collection and management of data (audience/operational/content data) at large scale to train relevant AI algorithms and the adoption of a "data-first" approach by media companies.

3.1.9.3 Emerj AI Sector Overviews (Better Software Group, 2019)

A series of reports on AI sector overviews have been published by AI market research and publishing company Emerj, aiming to explore important AI trends across industries, from automobile and financial to creative industry and beyond.[50] In the following, we summarise the insights offered by some of these reports, focusing on the news sector, streaming services, entertainment companies, and marketing.

Automated journalism

This article[51] presents different ways in which AI is being integrated in newsrooms. The main enhancements offered by AI include:

- ***Streamlining media workflows*** and automating tedious tasks to allow journalists to focus on creative tasks like reporting;
- Revealing insights by exploring correlations between data;
- Accelerating journalistic research;
- ***Countering disinformation***; and
- ***Automatic creation of news stories*** from raw data.

The article offers examples of AI tools used by popular news outlets like the Editor[52] app of NY allowing efficient editing of news articles through real-time semantic discovery of people, events, locations and dates mentioned in an article; the Perspective API[53] organising reader comments in articles based on their toxicity and thus allowing users to only interact with useful comments; the Juicer[54] app by BBC that monitors the RSS feeds of outlets around the world, extracts stories and assigns semantic tags to them, allowing journalists to quickly find the most

[50] Emerj AI Sector Overviews: https://emerj.com/ai-sector-overviews/
[51] C. Underwood, Automated Journalism – AI Applications at New York Times, Reuters, and Other Media Giants (2019): https://emerj.com/ai-sector-overviews/automated-journalism-applications/
[52] Editor, an experiment in publishing: https://nytlabs.com/projects/editor.html
[53] Perspective API: https://www.perspectiveapi.com/how-it-works/
[54] The Juicer API: https://bbcnewslabs.co.uk/projects/juicer/

relevant articles on a search topic; Washington Post's Heliograf[55] software for creating automatic real-time news reports for sport events and election results; Associated Press' NewsWhip[56] app for trend detection in social media and also analytics on user engagement with specific content or topics; the Wordsmith[57] platform used by Associated Press to write financial recaps; and Guardian's chatbot[58] that delivers personalised news content on demand based on real-time text exchange in Facebook Messenger.

AI for entertainment

This article[59] offers an overview of how top entertainment companies like Disney exploit AI to increase customer satisfaction and improve their revenue. Relevant AI application areas include:

- Development of ***interactive virtual characters***;
- ***Chatbots*** for famous cartoons of movie characters to promote relevant content;
- ***Improving image quality*** in older movies;
- ***Automatic creation of movie trailers*** tailored to different audience preferences;
- ***Audience analysis based on sensor data*** (e.g. from camera) and prediction of audience reactions to specific content;
- ***Sentiment analysis*** (from social media, product reviews, surveys, etc.) to measure audience engagement with content;
- ***Social network data analysis*** to assist marketing across multiple platforms.

In addition to these applications, Netflix is using AI not only for its recommender system but also to personalise its interface to keep user interest alive. For the latter, two applications are discussed in another similar article[60]: selection of appropriate still images for thousands of titles (considering actor prominence in film, image diversity, and maturity filters for offensive content) and also personalisation of thumbnail images for viewers, which has shown to benefit the promotion of less well-known titles.

AI for advertisement and marketing

The article[61] presents the most popular and some emerging cases of AI use in the advertisement and marketing sector. These include:

---

[55] Heliograf: https://www.wsj.com/articles/washington-post-to-cover-every-major-race-on-election-day-with-help-of-artificial-intelligence-1476871202
[56] NewsWhip https://www.newswhip.com/
[57] Wordsmith: https://automatedinsights.com/wordsmith/
[58] Guardian Chatbot: https://www.theguardian.com/help/insideguardian/2016/nov/07/introducing-the-guardian-chatbot
[59] E. A. Rayo, Artificial Intelligence at Disney, Viacom, and Other Entertainment Giants (2019): https://emerj.com/ai-sector-overviews/ai-at-disney-viacom-and-other-entertainment-giants/
[60] R. Owen, Artificial Intelligence at Netflix – Two Current Use-Cases (2022): https://emerj.com/ai-sector-overviews/artificial-intelligence-at-netflix/
[61] D. Faggella: Artificial Intelligence in Marketing and Advertising – 5 Examples of Real Traction (2019): https://emerj.com/ai-sector-overviews/artificial-intelligence-in-marketing-and-advertising-5-examples-of-real-traction/

- *Improved* ***product search*** (more accurate, faster, personalised). Future trends in this direction include search based on input images.
- *Improved* ***recommendation engines*** based on collected data about user behaviour, content engagement, sales, etc.
- ***Programmatic advertisement***, allowing automated buying and selling of ad inventory and real time campaign optimisations based on audience/user behaviour analysis.
- ***Marketing forecasting***, using marketing data like clicks, views, time-on-page, purchases, etc. to predict the success of a marketing campaign.
- ***Conversational e-commerce***, using chatbots that help users to select products and make purchases.
- ***Content generation***, including automated product descriptions or promotion articles.

3.1.9.4 AI Can and Will Revolutionize the Media Industry (Better Software Group, 2019)

This blog article[62] tries to map areas where AI can benefit the media sector, providing real-life examples of its impact. The following areas of application are examined:

- ***AI for automation of routine and mundane media workflows***, e.g. extraction of audio and creation of subtitles in video, automated content delivery, A/B testing of different product parameters to enable better decisions, etc. Examples of such automation include technologies that allow automated A/B testing for optimising Netflix's recommendation engine or automatic production of financial reports and news summaries by news organisations like Associate Press and Reuters.
- ***Archiving, metadata creation and improved content search*** based on video and audio analysis (recognition of faces, voices, objects, places etc.) Such techniques make content discovery easier and more accurate while also enabling automatic content moderation. In addition they facilitate content personalisation and thus content engagement.
- ***Content personalisation*** based on analysis of vast amounts of collected user data. AI can help find the balance between customisation and giving the user more of what they like on one hand and smart new content promotion on the other. Large media platforms like Facebook and Netflix but also Amazon have invested heavily on such technologies and their success is tightly connected to their personalisation services.
- ***Audience analytics*** capturing user behaviour and interactions with content and increasing insight on audience needs that allow media companies to understand their audience and effectively monetise their content. This feature allows better product/content recommendations, targeted advertisement at the right moment, and expansion of the audience base.

3.2 AI research and technology trends

In this subsection, we present a selection of roadmaps, surveys, and articles focusing on the main AI technology trends, going into 2022. Unlike the previous subsections, where the focus

[62] AI Can and Will Revolutionize the Media Industry, Better Software Group (2019): https://www.bsgroup.eu/blog/ai-will-revolutionize-media-industry/

was on AI application trends for the media, creative and entertainment industries, here we offer a more generic overview of AI technologies that have the potential to positively disrupt various industry sectors. Despite the lack of specific focus in the media industry, this general overview offers very helpful insights on future AI trends, allowing us to highlight the most promising technologies and identify opportunities for the take-up of these technologies by the extended media sector.

3.2.1 A 20-Year Community Roadmap for Artificial Intelligence Research in the US (CCC, AAAI, 2019)

The report titled "*A 20-Year Community Roadmap for Artificial Intelligence Research in the US*"[63] offers a roadmap for AI research and development in the US in the next two decades (2020-2040) and it is the result of coordinated effort by the Computing Community Consortium (CCC) and the Association for the Advancement of Artificial Intelligence (AAAI).

The roadmap presents through various examples the benefits that AI can bring in six areas: 1) boost health and quality of life, 2) provide lifelong education and training, 3) reinvent business innovation and competitiveness, 4) accelerate scientific discovery and technical innovation, 5) expand evidence-driven social opportunity and policy, and 6) transform national defense and security. The examples take the form of detailed vignettes that effectively and vividly describe how AI innovations could impact society and business. The format of this survey was the inspiration behind the structure of sections 5-10 of this deliverable (especially for the vignettes describing how media practitioners or users enjoy the envisaged benefits of AI).

Three major AI research priorities are identified based on the aforementioned societal drivers:

- ***Integrated intelligence***, including the combination of modular AI capabilities to create intelligent systems with broader capabilities; contextualisation of general capabilities to fit the needs of specific domains/individuals /organisations/roles through incorporation of existing knowledge and continuous adaptation; creation of open repositories of machine-understandable world knowledge and development of human cognition models to help AI systems understand how the world works and act accordingly.
- ***Meaningful interaction***, including enabling productive and fluent collaboration between humans and machines by endowing AI with capabilities such as reasoning, human mental state recognition, understanding of social norms, supporting complex teamwork, etc.; supporting human social online interactions by developing AI systems for deliberation (e.g. fact-checking), collaborative online content creation (e.g. collaborative arts or software) and social-tie formation (e.g. sophisticated human-machine hybrid technologies to enable new forms of human interaction); integration of diverse interaction channels and combination of different modalities (visual, verbal, emotional) to improve natural interaction of people with AI systems but also increase the accuracy and robustness of AI systems while also preserving user privacy; and

[63] Y. Gil and B. Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US (2019): https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/Community-Roadmap-for-AI-Research.pdf

responsible and trustworthy AI that explains its behaviour and actions, that can be corrected and improved directly by users, and that can act responsibly and ethically, accept responsibility, and get and uphold people's trust.

- ***Self-aware learning***, including trustworthy AI learning that ensures AI systems that know their own limitations and can provide explanations for their actions; learning expressive representations that go beyond correlation by using causal models and combining symbolic and numeric representations; durable AI systems based on efficient incorporation of prior knowledge and use of limited data (one-shot/few-shot learning); integration of AI and robotics for real-time learning of behaviours and actions.

For each of the aforementioned research priorities, a set of research areas has been identified while for each area a set of specific research directions is examined, identifying for each direction a set of milestones in the development of the specific AI functionality/capability/ resource. In Figure 7 below, we attempt to briefly summarise the AI research areas and directions highlighted by the roadmap.

Figure 7: A summary of research priorities, areas and challenges/directions for the 20-Year Community Roadmap for Artificial Intelligence Research in the US.

The roadmap also touches upon current AI challenges emphasising issues such as *AI ethics, privacy, security and vulnerability of AI systems, AI trust, availability of resources including datasets, compute, software, and finally human AI talent and skills*.

Although, the report provides an overview of these AI trends and challenges with a view to important societal needs like health, education, defense, social justice, business growth, etc., this analysis is very much relevant to the needs of the media and entertainment sector since it covers many different aspects of media-related AI applications/challenges (e.g. multimodal interaction with users, personalisation, self-aware learning, integrating intelligence from different fields, explainability, trustworthiness, ethics, etc.).

The report also includes a summary of major findings. AI is poised to have a profound impact in all sectors of society helping citizens in dire situations, educating them, and providing personalised services. To realise this potential, large interdisciplinary research teams are required supported by adequate resources (massive datasets, common architectures, shared software/hardware infrastructure) and facilitated by cross-fertilisation between different fields. AI research should be audacious and significantly more integrative and experimental while recognising what is necessary regarding the impact of AI in society. Solutions to major AI problems will come from the collaboration of the research community (that studies fundamental questions) and the industry (that has vast amounts of data, domain knowledge and compute). The demand and supply gap in AI talent will grow significantly in next decades, which necessitates AI education initiatives, collaboration between academia and industry, and action to increase diversity.

On the other hand, this rapid-deployment of AI systems raises a series of ethical questions and challenges with regard to security, privacy, ethical ramifications of AI-assisted decision making and content generation, oversight of and responsibility/accountability for AI decisions, fairness and transparency, which should be dealt with via a multidisciplinary approach that involves social sciences, humanities, and computing, preferably in the academic environment. The delivery of ethical next-gen AI systems that will bring major social and economic, technological and societal change will require significant strategic investments.

Based on these findings, the report offers recommendations along three main dimensions:

- *National AI infrastructure* (in the US but similarly in the EU), including the development of open AI platforms and resources (AI-ready data repositories, AI software and AI integration frameworks, AI testbeds); sustained community-driven challenges that will move the research forward in selected fields (like RoboCup[64]); national AI research centers conducting multidisciplinary research, developing open resources and providing AI training; and mission-driven AI labs, acting as living laboratories for AI development in areas of great societal impact, allowing collection of data and development of algorithms to tackle real-world problems (e.g. AI-ready homes, hospitals, science labs).

[64] The RoboCup initiative: https://www.robocup.org/objective

- *Core programs for AI research*, providing funding for basic AI research; application driven-research; interdisciplinary research; private-public partnerships; integration of AI research and education; diversity and inclusion; development of education curricula.
- *Training of a diverse AI workforce*. This includes development of AI curricula at all education levels; recruitment programs for advanced AI degrees; engagement of underrepresented and underprivileged groups to increase diversity; provision of incentives for interdisciplinary AI studies on subjects like AI ethics and policy or AI and the future of work; programs for training AI engineers and technicians; and retraining programs to convert technical personnel to AI engineers/ technicians.

Figure 8 summarises the analysis and suggestions of the roadmap.

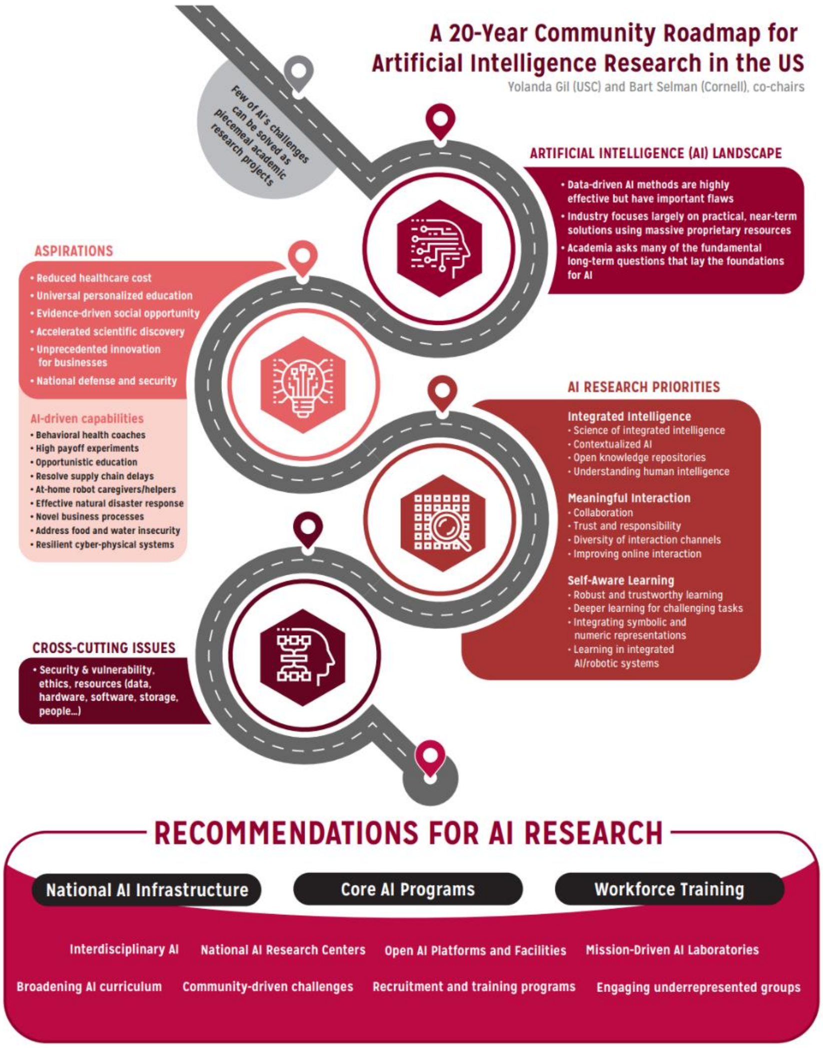


Figure 8: A 20-Year Community Roadmap for AI Research in the US.[65]

[65] Image source: Y. Gil and B. Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US (2019): https://cra.org/ccc/wp-content/uploads/sites/2/2019/08/Community-Roadmap-for-AI-Research.pdf

3.2.2 Artificial Intelligence Index Report 2021 (Stanford, 2021)

The AI Index 2021 Report[66] by the Human-Centered AI Institute of Stanford University aims to offer a thorough overview of all things AI, from research & development to education, ethical challenges and national strategies, providing both relevant data and useful insights.

The report highlights selected trends with regard to AI technologies and their potential adoption and impact on business, society and research. Below, we present those relevant to the media sector:

- **Generative AI**: AI systems based on generative technologies like GANs, transformers or VAEs will increasingly compose new media content (video, images, audio and text) of high quality, with a potentially tremendous range of applications, both useful (e.g. for the creative industry to improve film making and game development) and also harmful (e.g. deepfakes to spread disinformation). Effort is not only focused on new generative models but also on the detection of AI-generated content.
- **Industrialisation of computer vision**: Computer vision performance is starting to flatten on some of the largest benchmarks, suggesting harder ones are required. Increasingly large amounts of computational resources are invested to training relevant models at faster rates. At the same time, technologies like object-detection in video are maturing rapidly. All these indicate that AI-enabled computer vision will be further deployed by the industry in more real-life applications.
- **Natural Language Processing**: NLP has rapidly progressed over the last years, resulting in AI systems with significantly improved language capabilities that are being adopted more and more for different applications (e.g. conversational agents to assist users and consumers, automatic multi-lingual translation, automatic storytelling, robot journalism, etc.), starting to have a meaningful economic impact on the world. Progress in NLP has been so swift that technical advances have been outrunning relevant benchmarks.

The report also presents progress in various AI subfields like computer vision, language, and speech, highlighting technologies that have been showing accelerated progress and big potential, becoming more affordable and applicable to many more different areas:

- *Computer vision*: image classification, image generation, deepfake detection, pose estimation, semantic segmentation, embodied vision, activity recognition, object detection, face detection and recognition.
- *Language & speech*: language understanding, machine translation, language models (GPT-3), vision and language reasoning, speech recognition (speech transcription, speaker recognition).

[66] D. Zhang, S. Mishra, E. Brynjolfsson, J. Etchemendy, D. Ganguli, B. Grosz, T. Lyons, J. Manyika, J. C. Niebles, M. Sellitto, Y. Shoham, J. Clark, and R. Perrault, “The AI Index 2021 Annual Report,” AI Index Steering Committee, Human-Centered AI Institute, Stanford University, Stanford, CA (2021): https://aiindex.stanford.edu/wp-content/uploads/2021/11/2021-AI-Index-Report_Master.pdf

With regard to the most exciting trends, experts highlight the dominance of the ***Transformer architecture*** and the way advancements in NLP are having a profound effect on advancements in vision.

3.2.3 Artificial Intelligence: 7 trends to watch for in 2022 (Enterprisers Project, 2022)

This article[67] by The Enterprisers Project predicts that 2022 will be the year that AI will mature from experimental to essential across most industry sectors, shifting the focus to AI-enabled business transformation aiming to solve significant problems with AI-driven tools. The article identifies seven major AI trends, discussed below. As can be seen by the examples offered, these trends are increasingly applicable to the media sector.

- **Data wrangling**: The need for massive amounts of data to train AI models necessitates the development of new flexible data pipelines that can collect and harmonise structured and unstructured data from thousands of sources (e.g. user behaviour data, content engagement data, operational data, etc.), allowing real-time data processing.
- **Automated process discovery** (APD): Based on analysis of business data, AI can be used to automatically discover and map out an organisation's business processes, providing deep insights on the business. APD can thus facilitate robotic process automation (e.g. robot journalism), allowing some of the most mundane and tedious workplace tasks to be done automatically.
- **Intelligent supply chains**: AI can make supply chain management more effective. Relevant applications in the media include for example supply and demand planning of entertainment content across media platforms.
- **Customer-facing AI**, in the form of virtual agents and chatbots has been on the rise since the pandemic started and is expected to handle more complex cases of communication.
- **NLP**: NLP applications are expected to become mainstream, including creative writing and conversational agents.
- **AI to increase IT productivity**: AI will be used to improve management of IT systems, assisting humans and offering real-time actionable interventions. At the same time, generative AI could be employed for app development, increasing developers' productivity.
- **AI talent**: Companies will need to step up their game to train, recruit and retain AI talent. This necessitates initiatives that promote a culture of inclusion and life-long learning as well the development of partnerships across industries and organisations.

3.2.4 The 7 Biggest Artificial Intelligence (AI) Trends In 2022 (Forbes, 2021)

[67] S. Overby, Artificial Intelligence (AI): 7 trends to watch for in 2022 (2022): https://enterprisersproject.com/article/2022/1/artificial-intelligence-ai-7-trends-watch-2022

This article[68] by Forbes presents seven fields where important AI breakthroughs are expected:

- **Augmented workforce**: AI machines and tools can help the workforce to boost their own abilities and skills and work more efficiently, becoming a substantial part of everyday work.
- **Better language models**: The release of GPT-3[69], a language model with 175 billion parameters, has already changed the way NLP is being integrated in different industry applications. Its successor GPT-4 is expected to make natural conversation between humans and machines more plausible.
- **AI in cybersecurity**: AI can have an increasingly significant role in the fight against cybercrime in today's highly interconnected online environment by analysing vast amounts of network and other data to detect malicious intentions.
- **AI for the Metaverse**: Although, what the Metaverse will look like is not clear yet, it can be described as a fully digital world supporting all kinds of immersive experiences for the users. Obviously, AI has a tremendously important role to play in the Metaverse in many different ways: from creating the virtual worlds that humans will work or play in to developing sentient AI beings aiming to help us, entertain us or offer companion.
- **Low code & no-code AI**: One of the main reasons hindering the wide adoption of AI is the lack of skilled personnel to develop AI tools. Low code or no code AI aims to offer an easy-to-use interface that will allow the development of complex AI systems by plugging in various ready-to-use off-the-shelve AI components that will be then trained with a specific organisation's data. This will facilitate AI adoption and democratisation.
- **Autonomous vehicles**: AI is essential for autonomous vehicles. Significant advances are expected in this field in 2022 that may see the first self-driving car becoming reality.
- **Creative AI**: AI will be increasingly used to automate routine creative tasks such as creating headlines or graphics designing.

As in the previous case, the aforementioned trends are highly relevant for the media and creative industries allowing for example, automation of tedious tasks in newsrooms; automatic content translation or search in archives; more secure online communications and exchanges; realistic human virtual characters; easy creation of AI applications for the media; and new ways of creativity. Even autonomous vehicles could be used, e.g. to automatically record news or film footage in dangerous environments.

A relevant article[70] builds on the trends examined by Forbes and adds two additional areas of interest to the list: ***hyperautomation***, where robotic process automation technologies will be used to automate tedious processes or repetitive tasks, improving speed and efficiency and

[68] B. Marr, The 7 Biggest Artificial Intelligence (AI) Trends In 2022 (2021): https://www.forbes.com/sites/bernardmarr/2021/09/24/the-7-biggest-artificial-intelligence-ai-trends-in-2022/?sh=7a22025c2015
[69] T. Brown et al, Language Models are Few-Shot Learners (2020): https://arxiv.org/abs/2005.14165
[70] K. Vyas, Top 8 AI and ML Trends to Watch in 2022 (2021): https://www.itbusinessedge.com/it-management/top-ai-ml-trends-to-watch/

freeing the workforce to do more substantial work; ***NLP applications*** are expected to become mainstream, including creative writing, conversational agents, or even code generation.

3.2.5 Future Today Institute's 2021 Tech Trend Report - Artificial Intelligence (AI) (Future Today, 2021)

Future Today's report[71] explores future AI trends in different areas, including health, defense, enterprise etc., offering an in-depth insight and detailed list of emerging AI applications for these areas. For this deliverable, we examine the individual reports focusing on AI enterprise[72], consumers[73], and creative trends[74], which appear to be most relevant to the news and entertainment industry. In the following, we briefly list the major trends for each field.

Enterprise trends include AI trends that focus on specific AI research technologies that could have wider application in various industry sectors:

- ***Low-code or no-code machine learning*** will allow enterprises to build and deploy customised AI models with minimal AI coding skills through simple to use interfaces and ready to use ML modules such as recommender systems or image classification tools.
- ***Web-scale content analysis based on NLP*** that will facilitate analysis of large unstructured datasets, allowing efficient tagging and information detection, a feature that can be useful for investigative journalism (see Panama papers) or hate speech detection on social networks.
- ***Empathy and emotion simulation*** based on the detection of the user's or consumer's emotions through face, speech and text analysis.
- ***Artificial emotional intelligence*** aiming to teach machines (e.g. virtual AI assistants) to convincingly exhibit human emotion.
- ***AI at the edge*** will allow processing of data closer to the source (e.g. user's mobile phone) thus increasing privacy and speed.
- ***AI chips*** dedicated to specific AI tasks with high-computational load (e.g. training NLP or object detection models) will provide faster and more secure processing thus accelerating the commercialisation of AI applications.
- ***Detection of fake media content,*** including deepfakes, fake posts/reviews, falsified documents, etc.
- ***NLP for ESGs***, i.e. to identify, tag, and sort documentation from various sources about a company's environmental, social, and governance (ESG) reputation.

---

[71] Future Today Institute, 2021 Tech Trend Report - Artificial Intelligence (2021): https://futuretodayinstitute.com/trends/

[72] Future Today Institute, 2021 Tech Trend Report - Artificial Intelligence - Enterprise (2021): https://olc.worldbank.org/system/files/2.%20Enterprise.pdf

[73] Future Today Institute, 2021 Tech Trend Report - Artificial Intelligence - Consumer (2021): https://olc.worldbank.org/system/files/4.%20Consumer%20and%20Research.pdf

[74] Future Today Institute, 2021 Tech Trend Report - Artificial Intelligence - Creative (2021): https://olc.worldbank.org/system/files/5.%20Talent%20and%20Creative.pdf

- ***Intelligent OCR***, to recognise both text and context in written documents.
- ***Robotic Process Automation***, to automate tedious tasks in the workplace, increasing productivity and creativity.
- ***Massive Translation systems*** that can translate content in multiple languages and dialects in real-time.

The report also highlights important AI risks and challenges including ***insurance liability*** for AI decisions and actions; ***manipulation of AI*** *for competitive advantage* (e.g. big-tech companies manipulating their search algorithms to prioritise profitable results); development of ***marketplaces*** *to buy and sell AI algorithms*; or development of ***software that is viable over time*** and can adapt to changing environments and resources.

Consumer trends includes AI applications such as:

- ***Ubiquitous virtual assistants*** that use semantical and natural language processing as well as collected user data to anticipate what we may need, answer questions and provide information, or assist in everyday tasks. The news and entertainment industry can harness digital assistants to both collect data but also deliver information.
- ***Deepfakes for fun*** applications, allowing users to manipulate their face in creative ways and entertain themselves or their friends.
- ***Personal digital twins*** that can learn from the user and represent them online, e.g. in the Metaverse and social media or other more professional settings.

Creative trends include AI trends that can help boost creativity, a feature especially important in the media sector. Such trends include:

- ***Assisted creativity***, aiming to enhance the creative process. Algorithms like GANs are used to create new music, images, 3D worlds, allowing artists and the general public to express their creativity in new ways.
- ***Generative algorithms for automated content production****,* including deepfake videos or music performances for the entertainment industry.
- ***Generation of virtual environments*** from short videos based on GANs that can be used in the film and game industry to reduce production costs.
- ***Automated versioning*** to develop different versions of the same media content, to reach a wider or different audience or produce content at scale.
- ***Automatic voice cloning and dubbing*** can have numerous applications in the media and entertainment sector, e.g. for making actors or newscasters speak fluently in different languages.
- ***Automatic ambient noise dubbing*** will make it easier to generate ambient sounds and thus automatically generate videos or storylines.

AI research trends: Apart from AI trends for various sectors, the Future Today Institute's report provides a very detailed list of research trends, summarising what AI technologies seem to hold an increased potential for the future. Below, we present the list of the most important technologies without going into details:

- *Machine reading comprehension* to read, infer meaning, and answer while sifting through enormous datasets;
- *Federated learning* for ML at the edge;
- *NLP models* for automatic text and story generation;
- *Vokenisation* to infer context from images. This technique maps language "tokens" such as words to related images or "vokens";
- *Machine image completion* to autocomplete and enhance images;
- *Predictive models* such as GANs using a single image to predict, for example, what happens next in a video;
- *Real-time ML* to collect and interpret data, incorporate context, and learn in real-time;
- *Automated ML* that allows matching raw data with models to obtain useful information without requiring AI experts and time-consuming processes;
- *Hybrid human-computer vision* combining humans and AI for greater accuracy in computer vision tasks;
- *Neuro-Symbolic AI* to combine logic and learning thus creating systems that may not need humans to tag the data and train a model;
- *Reinforcement learning* for developing AI agents that can learn multiple tasks based on reward/punishment;
- *Continuous learning* to facilitate autonomous and incremental skill building development of AI agents.

3.2.6 State of AI report 2021 (stateof.ai, 2021)

This annual report[75] presents important AI technology breakthroughs that hold special promise for the future. In the 2021 version, the authors single out the following AI technology trends that can have an application to the media sector:

- ***Vision Transformers***: the transformer architecture has expanded beyond NLP and is expected to provide state-of-the-art results in computer vision tasks like image classification, scene segmentation and object detection. This technology can also significantly benefit research in other major AI application fields like audio (e.g. speech recognition) or 3D point clouds (e.g. 3D object classification or scene segmentation). The report predicts that Transformers will "*replace recurrent networks to learn world models with which RL agents [will] surpass human performance in large and rich game environments*";
- ***Self-supervised learning***: self-supervised learning on large datasets is expected to take over computer vision tasks, following its success in NLP research;
- ***Reinforcement learning***: RL techniques are used to develop AI agents for increasingly complex games. The main challenge lies in their ability to generalise. RL agents trained

[75] N. Benaich and I. Hogarth, State of AI Report 2021: https://www.stateof.ai/

by DeepMind were shown to be able to generalise well in new games without additional training, exhibiting behaviours such as experimentation and cooperation. This ability holds great potential for the use of RL in developing AI agents for different applications;

- ***Generative Spoken Language Modelling***: This technique learns speech representations from raw audio without requiring labels or text. This "textless NLP" will improve speech generation for rarer languages (thanks to the large number of audio data from local podcasts or radio shows that can compensate for the limited textual information available online) but will also enhance expressiveness of generated speech by exploiting emotion and other nuances recorded in audio;
- ***Diffusion models***: Diffusion models are already beating GANs in tasks such as image generation, audio synthesis, or music generation, providing a promising technology for new content generation;
- ***Prompt-based learning for NLP models***: Use of pre-trained NLP models for new tasks is shifting from a "*pre-train, fine-tune*" process (where we try to adapt the model to new tasks via objective engineering) to a "*pre-train, prompt and predict*" approach[76] (where we try to adapt the task to the model, i.e. to reformulate it to look more like the tasks solved in the original training, through a *textual prompt*)[77]. Prompting is important to few-shot or even zero-shot learning; however, prompt selection is quite challenging and can lead to performance degradation. To this end, prompt learning is proposed.
- ***Graph Neural Networks***: GNNs are already becoming one of the hottest AI topics with applications in computer vision, text analysis, physical system modelling, medicine, etc.

3.2.7 Ultimate Guide to the State of AI Technology in 2022 (AIMultiple, 2022)

This short guide[78] provides a brief look on the status of AI technology, focusing on three levels: AI algorithms, computing technology to run the algorithms, and applications to different domains. It makes a special reference to the **computing power necessary to train AI systems**, referencing a relevant analysis published by OpenAI[79], which shows that since 2012 the required computing power for training AI algorithms has been increasing in a considerably faster rate (3.4-month doubling period) than Moore's law predicts. These increased computational needs have the potential to limit future AI technology advancements since advances in computing

[76] P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, https://arxiv.org/abs/2107.13586

[77] *"For example, when recognizing the emotion of a social media post, "I missed the bus today.", we may continue with a prompt "I felt so ", and ask the LM to fill the blank with an emotion-bearing word. Or if we choose the prompt "English: I missed the bus today. French:"), an LM may be able to fill in the blank with a French translation. In this way, by selecting the appropriate prompts we can manipulate the model behavior so that the pre-trained LM itself can be used to predict the desired output, sometimes even without any additional task-specific training"* (example taken from P. Liu, W. Yuan, J. Fu, Z. Jiang, H. Hayashi, G. Neubig, Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing, https://arxiv.org/abs/2107.13586)

[78] C. Dilmegani, Ultimate Guide to the State of AI Technology in 2022 (2022): https://research.aimultiple.com/ai-technology/

[79] D. Amodei and D. Hernandez, AI and Compute (2018): https://openai.com/blog/ai-and-compute/

power do not seem to be able to follow the exponential growth of AI needs. To increase computing power aiming to serve the needs of AI technology, two main trends are identified.

The first is the development of ***AI-enabled chips*** that will be optimised for running machine-learning workloads, e.g. training AI models for dedicated complex computer vision or NLP tasks. Big investments on relevant start-ups highlight the potential of this technology. According to Allied Market research[80], the global AI chip market size is expected to grow from $8.02 billion in 2020 to $194.90 billion in 2030, with a CAGR of 37.41%.

In addition to such AI chips, the article foresees that new computing technologies like ***quantum computing*** can become a game changer that will allow AI to maintain its rapid growth rate. The International Data Corporation in its first forecast for the global quantum computing market[81], projected that investments in quantum computing will reach $16.4 billion in 2027, with a CAGR of 11.3% in 2021-2027.

With regard to AI trends, the article points out the potential of technologies such as ***Reinforcement Learning, Transfer learning, and Self-Supervised Learning*** while also drawing attention to three prominent areas of application, highly relevant to the media sector: *computer vision, natural language processing, and recommendation systems*. In a complementary guide[82], additional trends and application domains are identified. These include: ***Explainable AI*** to help companies and users understand how AI models work; ***fusion of AI and cloud*** *technologies* aiming to allow AI models to collect data from the cloud, self-train, and then apply the newly acquired insights into the cloud where they can be used by other models; ***AI for XR*** *experiences* supporting touch, smell and taste; and ***merging of AI and IoT*** *technologies* to allow devices to learn from their data and make better decisions.

3.3 Summary of AI applications, AI technologies and AI challenges for the media sector

In the previous subsections, we tried to offer an in-depth overview of the different AI applications and AI research trends that can have an impact on the media sector, by presenting the findings and insights from a selected number of roadmaps, surveys, white papers, articles, etc. This presentation revealed a landscape where the opportunities for the use of AI are enormous while the variety of tasks across the media supply chain that AI can improve, assist, automate, expand or create is limitless. AI can have a truly transformative influence on the media sector, reinventing the business model of media organisations, establishing new ways of work and increasing the productivity and creativity of the workforce, and finally transforming and enhancing the user experience across platforms.

---

[80] A. Savekar and S. Sachan, Artificial Intelligence Chip Market Report 2021 (2021): https://www.alliedmarketresearch.com/artificial-intelligence-chip-market

[81] International Data Corporation, Worldwide Quantum Computing Forecast, 2020-2027: An Imminent Disruption for the Next Decade (2021): https://www.idc.com/getdoc.jsp?containerId=prUS48414121

[82] C. Dilmegani, Future of AI according to top AI experts: In-Depth Guide (2022): https://research.aimultiple.com/future-of-ai/

In the following, we will attempt to summarise the most prominent AI-enabled applications for the media sector. Figure 9 presents the relevant applications per industry, focusing on news, film/TV/streaming, music, games, social media, advertising, and publishing. As can been seen, each sector has its own unique needs (e.g. fact-checking for newsrooms or AI-based casting for films) but most of the applications featured in Figure 9 aim to satisfy similar needs, e.g. the need for content personalisation or automated content creation (whether this is music, films, ads, books or games), better recommender systems, enhanced understanding of users, etc.

News

- Robot journalism/automated reporting
- Fact-checking
- Content verification (deepfake detection)
- Content archiving & search (video, image, text, etc.)
- Automated transciptions
- Multi-lingual content translation
- Social media analysis for trend detection
- AI assistants for journalists (for editing/writing/visuals)
- Augmentedj journalism (using drones, wearables, voice, VR for novel content creation and delivery)
- Recommender systems
- Personalised content creation & delivery
- Chatbots to assist subscribers/audience
- Content moderation (e.g. comments on articles)
- Compliance with copyright standards
- Audience analysis
- Sentiment analysis of user content
- Hyper-targeted advertising
- Forecasting (subscriptions, trends, sales, content monetisation)

Film/TV/Streaming

- Personalised programming
- Content recommendation
- Content personalisation (e.g. personalised movie trailers, user-driven storylines, interfaces etc.)
- Automated content generation (script, voice, video, CGI, deepfakes, trailers, video highlights, live commentary, captioning, etc.)
- Multi-lingual translation
- Content enhancement (e.g. film restoration)
- AI-based casting
- VR-enabled user experiences
- Sentiment analysis (from social media, product reviews, surveys, etc. but also using in-room sensors) to measure audience engagement with specific content
- Chatbots to assist subscribers/audience
- Audience analysis
- Churn prediction
- Dynamic product placement and advertising
- Programmatic ad buying
- Hyper-targeted marketing
- Forecasting (sales, subscriptions, trends, audience engagement, residual payments, content monetisation)

Music

- Automated content creation (music, singing voices, sounds, music clips etc.)
- AI creativity assistants for music creators and users
- AI-enabled real-time feedback for artists/creators
- Audio indexing and search
- Demixing
- Content recommendation
- Content personalisation (e.g. music matching our mood)
- Chatbots to assist subscribers/audience
- Customer base segmentation
- Forecasting (sales, music trends, audience engagement, residual payments to talent, content monetisation)

Games

- Generative game design
- Procedural content generation (graphics, music, etc.)
- Sentient AI agents / virtual characters
- Player profiling
- Personalised games dynamically adapted to players
- Personalised marketing
- Multi-lingual translation
- VR-enabled user experiences

Social media

- Reccomendation engines
- Content personalisation
- Enhancde content search
- Multi-lingual translation
- Automatic ad placement
- Chatbots to assist users
- Trend detection
- Opinion mining
- Content moderation
- Monetisation of user generated content
- Forecasting (ad sales, user engagement with content, content monetisation, trends, revenue)

Advertisement

- Emotion-based advertising
- Hyper-targeted advertising
- Programmatic ad buying
- Customer base segmentation
- Multi-lingual translation of ad content
- Chatbots to assist consumers/users
- Market forecasting (sales, campaign success, content engagement, etc.)
- Automatic content generation (ad scripts, ad videos, ad graphics, promotional material, etc.)

Publishing

- Automatic content generation (book summaries, user review summaries, graphics, imagery for children's books, voice for audio book etc.)
- Content editing
- Content indexing & search
- Content personalisation (e.g. personalised e-books or audio books)
- Content recommendation
- Multimodal interactive experiences (e.g. for e-books)
- Multi-lingual translation of content
- AI assistants to support publishers, editors, graphic designers
- Improved accessibility for impaired users
- Audience segmentation
- Detection of trends in content consumption/production
- Identification of users /prosumers (authors, fans, influencers, trend-setters etc.) and monitoring of community dynamics
- Co-creation and distributed mentoing in fanfic communities
- Copyright management
- Forecasting (sales, trends, content appeal, etc.)

Figure 9: A summary of AI applications for the media and entertainment industry.

AI applications for the media sector

The list below summarises applications of AI that are already having or can have a significant impact in most media industry sectors, addressing common needs and shared aspirations about efficient workplace automation, enhanced content, enhanced user experiences, better understanding of user needs, better marketing[83]:

Smart recommender systems that will analyse vast amounts of user and content data (also combining them with trend analysis or info about current state of affairs in the neighborhood and the world) to recommend content that matches user preferences or current mood or even needs that they did not know they had.

Content personalisation that will allow media companies to offer content not to audiences of billions but to billions of individual users with their own unique preferences, experiences, needs and moods. The trend is clearly more and more personalisation, going beyond simple user preferences derived from past actions and behaviours and adopting more elaborated approaches that also take in mind what happens in the moment to the user herself and in her environment. Personalisation involves content presentation (e.g. writing style of news digests), interaction with content (e.g. user-driven film storylines), and personalisation of content itself (e.g. music that matches our mood, game visuals that match our aesthetics, personalised movie trailers that will excite us.

Automated content creation (text, image, video, audio, VR), aiming to improve productivity and enhance creativity of media professionals but also to offer new ways of creative expression to the users. The potential applications of this technology are unlimited: robot journalism from automatic headlines to article writing, procedural content creation for games, creation of new music, deepfakes for the film industry, automated TV commentary for sports events, graphics for ads, book/film/podcast summaries but also creation of new art (image, video, music) by talent and simple users.

Enhanced content search for media professionals and users to allow efficient real-time retrieval of relevant content and thus improve exploitation and monetisation prospects. This will require automatic metadata tagging and extraction based on advanced video, audio and text analysis of content (detection and recognition of faces, voices, objects, places, dates etc.) that will make content discovery easy and accurate while also enabling other functionalities like automatic content moderation or content recommendations. Search will not be limited to text queries but will support voiced questions, audio, images, sketches etc. Again, the applications are numerous: users find the online content they search for with less effort, journalists can search large audiovisual archives to find historical information or social media and other outlets' websites, music composers can find audio excerpts to inspire them, book publishers can search the vast volume of titles produced every year, etc. Efficient content indexing and search means that the content can be more easily monetised.

[83] The icons used in this and the next page are from flaticon.com and vecteezy.com

Audience analysis aiming to capture user behaviour and interactions with content and increase insights on audience needs. By understanding their audience and what makes them happy, media companies can more effectively monetise their content through personalisation of services/content as well as high-accuracy ad targeting. Audience analysis is based on large amounts of data, including user behaviour but also user generated content like posts, reviews or likes. A new promising direction in audience analysis is the use of sensors that can accurately and in real-time record the emotional reactions of users to content, allowing in-depth analysis of user interest and engagement (e.g., Which parts of the film were the most entertaining or emotionally moving? What storyline plots created bigger engagement?).

Social media analysis to detect trends about what is interesting right now or what people think about a specific issue. Trend detection allows media companies and professionals to react in real-time to what is happening to the world or their users and adapt their content and services accordingly.

AI assistants that can help media professionals do their job more efficiently but also assist the audience when using media services or trying to access content. In the first case, AI assistants support professionals in everyday tasks (e.g. an AI assistant that helps journalists by suggesting story topics based on trend analysis of social media, by searching the archives to find statements of a politician on a specific subject, by suggesting visuals to accompany stories, by validating content veracity, by answering questions etc.) or even replace them in doing tedious and boring tasks that limit creativity (e.g. an assistant that produces daily financial news summaries allowing the journalist to focus their efforts into meaningful commentary on current financial situation). In the second case, AI assistants in the form of chatbots or talking digital characters are increasingly used to help the users to find the content they want, to answer user questions, or help users play a game. Advances in NLP promise to make this kind of communication much more natural and effective.

Multi-lingual translation to help the media distribute and promote their content and services more efficiently, reaching audiences around the globe. This will allow content to be offered in the user's native language, users to communicate with each other without language barriers and professionals to search, analyze and use content from different languages. This development will also be a decisive step towards democratising access to content and media services as well as participation in the digital online world.

Forecasting to facilitate accurate predictions with regard to a variety of subjects such as content engagement, emerging trends, user behaviour, sales, subscriptions, content monetisation, churn, ad revenue, etc. Predictive analytics are essential for everyday tasks (e.g. organisation of marketing campaign and TV programming) but also for designing long-term strategies.

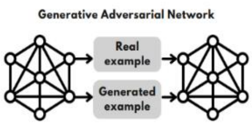
Hyper-targeted advertisement and programmatic ad buying. Advertisement is essential for the survival of the media industry both for promoting their own content and services aiming to create revenue from their content but also as the promoters of third-party goods, which results in revenue from brands that wish to advertise their products through the media company's channels. By analyzing vast amounts of user, content, operational and other (e.g. societal, financial, environmental) data, media companies but also those who advertise in them can target users in real-time with highly effective personalised ads that will make them loyal subscribers or customers. In addition, real-time analysis of audience dynamics and of available ad space across multiple channels can enable automated buying and selling of ad inventory and real-time marketing campaign optimisations.

Compliance with copyright standards. Copyright in media and creative industries is a huge issue that can cause a lot of headaches to media professionals and have serious legal implications. AI can transform this area by automatically analysing relevant legal documents or online copyright information thus saving thousands of hours for the legal department but also allowing media professionals to quickly identify IPR issues of online content.

AI technology trends for the media sector

To transform the media industry through the aforementioned applications well-known AI and ML technologies such NLP, RL, GANs etc. are already employed to harness the data and deliver the envisaged functionalities. In the following, we summarise a list of AI technologies that hold the greatest potential to realise the media's vision for AI. The list is obviously not exhaustive (and may be biased towards AI4Media's research activities and use cases) but reflects efficiently the most important current and future trends identified by AI experts and industry stakeholders in the surveys, reports and articles presented in the previous subsections[84].

Reinforcement learning: RL techniques are currently used to train AI agents that can play increasingly complex games, beating human champions. Their ability to learn goal-oriented behaviours through reward and punishment strategies will be increasingly used to develop AI agents for different applications, including AI agents that assist and learn from humans and their environment (for office automation tasks but also as companions in the digital world), smart recommenders that learn from user interaction with content and recommend content that will maximise user engagement and satisfaction or chatbots that self-improve.



Generative AI: Generative AI is used to automatically create new content by utilising existing content. Technologies like GANs, transformers, VAEs or diffusion models will be increasingly used by the media industry to create new high quality text, image, video and audio. The range of potential applications is without limits: deepfakes for the film industry, music composition, game assets, script creation, film preservation etc.

Transformers: After revolutionising natural language processing thanks to their computational efficiency, transformers are set to also dominate the computer vision

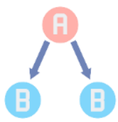
[84] The icons used in this and the next page are from flaticon.com and vecteezy.com

field by allowing training at large-scale with vast image databases to enhance the performance of image classification algorithms. This trend will accelerate progress in visual search algorithms that are increasingly important for efficient content indexing, search and recommendation.

Few shot or no shots learning: Big breakthroughs in AI have been enabled mainly by the development of models trained with massive amounts of data. However, it is not always plausible to have such amounts of data. Few shot learning techniques rely on limited available examples to learn models, which could come handy e.g. in the case of building NLP models for rare languages where limited data is available or voice cloning from few audio samples.

Emotion AI: Emotion AI includes technologies that learn and recognise human emotions. This knowledge can be used to help machines recognise a human's emotional state and adapt their behaviour to human emotion in addition to human action but also to teach machines to exhibit human emotion. The aim is to make human-machine communication natural, imitating the way humans would communicate with each other, but also to use knowledge about human emotion to offer better personalised services. The applications are numerous: chatbots, sentient virtual characters, targeted advertisement, understanding the audience, content recommendation or delivery based on user mood, etc.

NLP and multi-lingual NLP: NLP has witnessed a true revolution during the last few years with large language models like GPT-3. NLP is expected to become increasingly mainstream in the media business through applications such as conversational agents and virtual characters, creative writing, robot journalism, interactive storytelling, voice search for image/video/audio, sentiment analysis in social media, voice dubbing, or multi-lingual translation. Multi-lingual translation in particular will be a real breakthrough, breaking language barriers and allowing, on one hand, content creators to reach new audiences worldwide but also to exploit creatively the wealth of content available online (which is currently out of reach because it is in other languages) and, on the other hand, helping audiences and users to communicate freely and benefit equally (and more democratically) by the content created all over the world.



Causal AI: The *20-Year Community Roadmap for AI Research* makes several mentions to this technology which helps to move the needle beyond correlations to the identification of causal relationships. Causality is still an unfulfilled aspiration of AI which however receives increasing interest. Causal AI can have a significant impact on the media, since it will allow understanding the why and examining the what if: why users like some content more than other, why the recommendation algorithm provided this content, what would happen if a different film casting decision would be made, etc. The applications are many: explainability of AI decisions, better content personalisation, prediction of trends, prediction of content engagement, better content recommendations, more natural interaction with virtual characters (that understand why we act in a specific way and respond accordingly) etc.

AI at the edge: This is a highly emerging area of research in AI that aims to facilitate processing of data closer to the end-user device (e.g. in the user's mobile phone). This will significantly enhance data privacy and processing speed. Interesting applications include learning models (e.g. for content personalisation or development of AI agents) that learn continuously from the user by collecting sensor data from the home environment as well as real-time user feedback without allowing the collected data to leave the user's device.

Quantum computing: This emerging technology still has a long way to go. Its potential however would be game-changing for the media industry. These unforeseen capabilities of processing power will facilitate and accelerate machine learning using huge volumes of data, allowing for example social media analysis at scale, development of large multilingual language models and video game characters that behave hyper-realistically.

AI that learns like humans: Deep neural networks are inspired by biological neural networks of animals. Despite revolutionising AI and bringing major advances in machine cognition, DNNs are still a long way from enabling human-like intelligence especially with regard to humans' ability to easily generalise and learn new tasks with minimum training from infancy. An emerging trend in AI focuses on trying to understand the human cognitive system and to imitate human brain capabilities. Bio-inspired learning is expected to transform the way we interact with machines, allowing the development of AI agents with human-like intelligence and capabilities. Besides obvious applications in gaming and the Metaverse, such human like capabilities will have a profound impact in many different applications from search of content to AI-enabled creativity.

Other interesting trends include the **fusion of AI and cloud** that will allow collection of data from the cloud, self-training of models with this data, and availability of learnt knowledge in the cloud for others to exploit. Also, **convergence of AI with blockchain technology** has the potential to improve training data annotation and model training.

Challenges of AI adoption in the media sector

The previous analysis revealed the vast potential of AI to bring positive change to the media industry sectors. However, the works examined in the previous subsections highlighted that with high potential also come significant challenges and risks. We briefly summarise the main challenges identified below[85]:

AI explainability: Currently, AI systems are mostly black boxes without being able to explain why they recommended a product or predicted the success of a film or made a moderation decision. In order to fully adopt and trust such AI systems, media professionals but also users need to understand how such systems work. Explainable AI aims to do just that, increasing transparency and increasing trust and adoption of AI-enabled applications.

[85] The icons used in this and the next page are from flaticon.com and vecteezy.com

AI robustness: performance of AI algorithms may be hindered by many reasons, including malicious adversarial attacks but also poor performance when dealing with data different from those they were trained with. To ensure robustness, tools that help fortify AI models against attacks, predict new types of attacks, and ensure that the models perform as well in the real-world as they do in a sandbox are increasingly necessary.

AI bias and AI fairness: AI systems often exhibit bias against specific groups of people, including racial bias, gender bias, etc. due, for example, to prejudiced hypotheses made when designing the models or due to problems of diversity and representation in training data. AI bias can lead to bad business decisions or discriminate against groups of users. A prominent example for the media sector is bias that may be embedded in large language models. Such models are trained with swaths of Internet data, which are by definition produced in the biggest or richest countries, in languages with higher linguistic footprint, and by communities with large representation, or mainly by men[86], thus resulting in models that fail to capture changing social norms or the culture of minorities and underrepresented groups and which will eventually discriminate against such groups or produce language that is not attuned to changing social norms[87,88]. The gigantic volume of data also makes it hard to audit such models for embedded bias. To address this problem, new techniques and new frameworks have been proposed aiming to enhance AI-fairness and minimise bias.

Privacy concerns: AI applications like recommender systems or content personalisation are based on the collection of vast amounts of data about user's preferences, behaviours, actions, as well as user generated content. Obviously, this creates a lot of concerns about privacy and how this data may be used. To address such concerns, the EU has proposed regulations like the GDPR while companies are starting to explore solutions that will enhance the privacy of the users and their data.

Data for AI training: Many of the examined reports highlighted the need for large volumes of real and high-quality data for training AI models for the media industry.

AI skills: One of the reasons hindering the adoption of AI in the media industry is the lack of relevant skills by media professionals and challenges in recruiting AI experts. To overcome this obstacle AI training and education are necessary for media professionals as well as raising awareness about AI and its potential across an

---

[86] Such models usually get trained with data scrapped by sources like Wikipedia or Reddit where women are significantly under-represented. According to this Guardian article, women are less than 20% of the contributors of Wikipedia; according to Statista, women represent only 37% of Reddit users worldwide.

[87] K. Hao, MIT Technology Review, "We read the paper that forced Timnit Gebru out of Google. Here's what it says" (2020): https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

[88] E. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? In. Conf. on Fairness, Accountability, and Transparency (FAccT '21), March 2021.

organisation. Collaboration of the media industry with academia/research but also with other media organisations or industries on AI topics of common interest would also be beneficial.

***AI strategy**:* Another issue concerning the experts is that still many media companies do not have a clear AI strategy that will allow them to efficiently adopt AI in the workplace, recruit or train staff, make investments in specific technologies, pursuit useful collaborations and fully exploit AI’s potential for the media.

4 Online survey on AI for the Media Industry

One of the tools used to deliver this roadmap on AI for the media industry was a public survey addressed to both AI researchers working on multimedia AI but also to people working in the media industry or whose work is closely related to this industry (e.g. researchers studying the media, media regulators, people working in relevant NGOs, etc.) The survey aimed to collect their ***opinions on the benefits, risks, technological trends and challenges of AI use in the media industry*** as well as their experience on AI strategies and AI skills in media organisations, their insights on the most promising ways to facilitate AI adoption and knowledge transfer and, finally, their perceptions about ethical use of AI.

The survey was launched online and was disseminated through various channels that targeted the two aforementioned target groups, resulting in the collection of 150 responses from AI researchers and media professionals from 26 countries in Europe and beyond.

In addition to this survey that focused on AI for the media industry, another small-scale survey was launched to explore media AI and its impact on society and democracy, thus attempting to explore the second part of the project title, i.e. AI in the service of society and democracy. This short survey was internal, addressed only to AI4Media partners, and aimed to collect their opinions on the ***benefits and risks of media AI for the society and democracy*** as well as to record their views with regard to potential policies for the ethical use of media AI, aiming to safeguard fundamental human rights. 31 responses to this survey have been received from media professionals and AI researchers that are part of the project consortium.

In the following subsections, we present the results of the analysis of the two surveys, offering commentary and insights. The survey questionnaires are presented in the Appendix (section 15).

4.1 AI4Media online survey on AI for the Media Sector

An ***anonymous*** online survey was launched in December 2021 by the AI4Media project, aiming to collect the opinions of the AI research community and representatives of the broader media industry with regard to the most important trends, benefits, challenges, risks, facilitators, and ethics for the use of AI in the media sector. The survey included two slightly different versions of the same structured questionnaire, addressed to the AI research community and the media professionals' community, respectively (see Appendix, sections 15.1 and 15.2, respectively). The majority of the questions were the same but there were some additional questions addressed specifically either to the first or second community. The survey could be accessed through the AI4Media website[89] (see Figure 10) and was widely disseminated through different channels, including email, posts in relevant industrial/research/ EU forums, social media posts/campaign (see Figure 11 and Figure 12). In total, 150 people completed the survey, 98 from the AI research community and 52 from the media industry or organisations related to the media. In the following, we briefly describe the target groups and content of each questionnaire.

---

[89] AI4Media 1st Online Survey on AI Technologies and Applications for the Media Sector (2021): https://www.ai4media.eu/1st-survey-on-ai-for-media-2021/

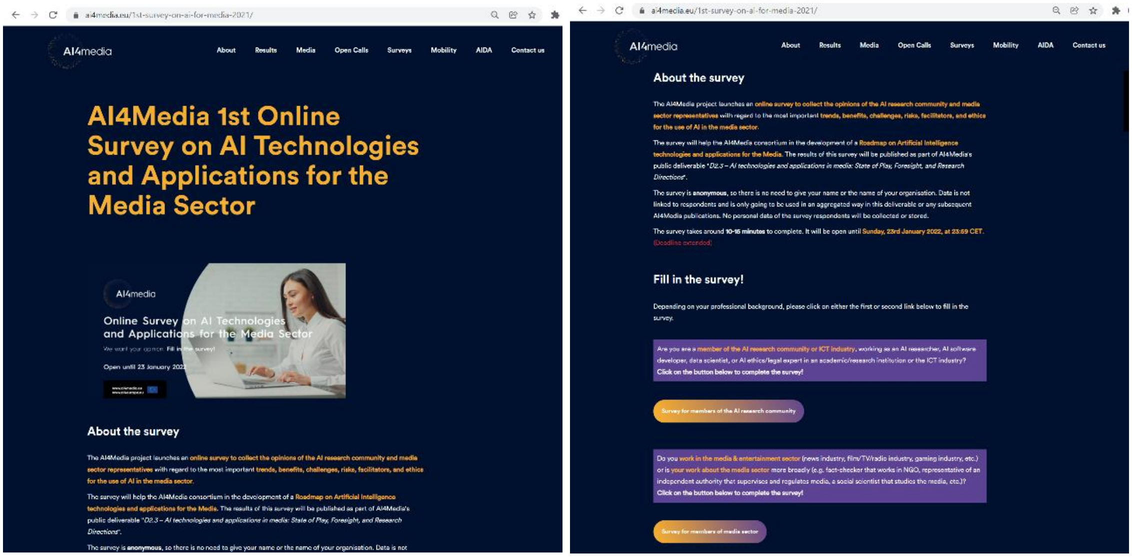


Figure 10: Webpage of the AI4Media online survey on AI for the Media Industry.

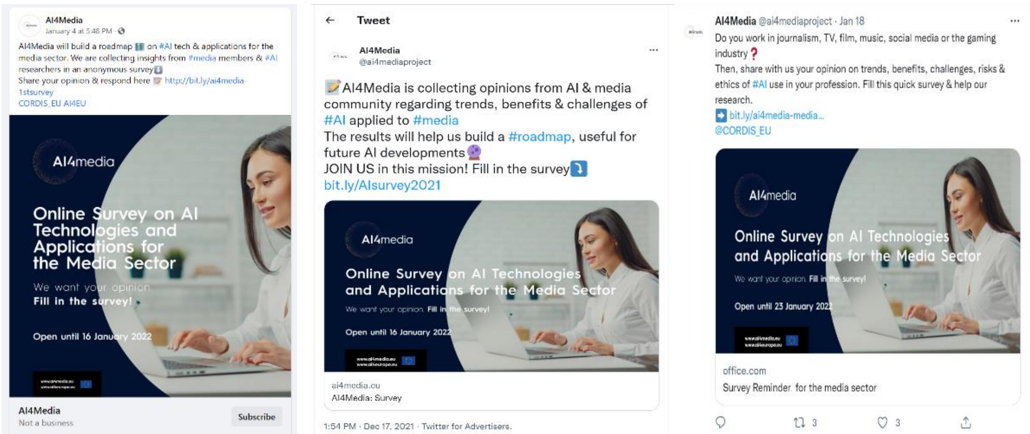


Figure 11: Social media posts disseminating the AI4Media online survey.

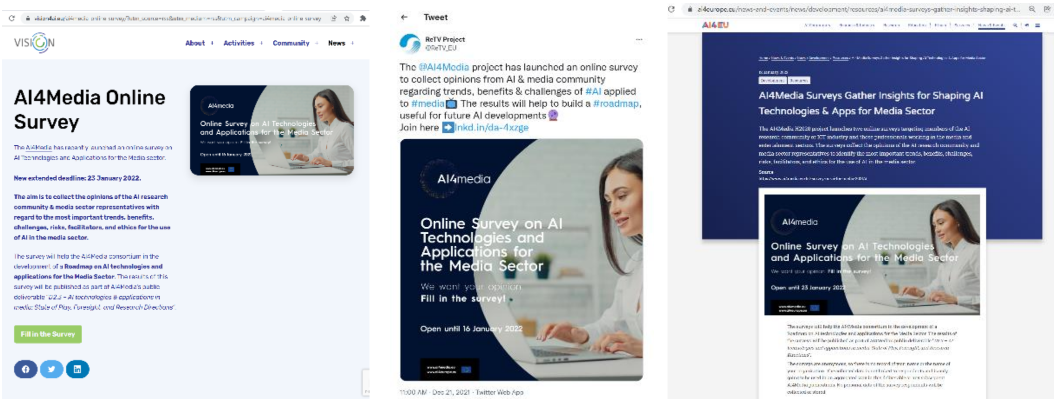


Figure 12: Dissemination of AI4Media online survey by other EU projects.

Questionnaire for the AI research community *(AIResearchQuest)*: This questionnaire was targeted at AI researchers and developers working in academia, research or ICT industry. The survey included the following sections: (1) Professional background of survey respondents; (2) Future AI technology trends for the media sector; (3) AI benefits for the media sector; (4) AI risks for the media sector; (5) AI challenges for the media sector; (6) Adoption of AI solutions by media sector; (7) AI ethics & AI regulation. In total, 98 members of the AI research community completed this questionnaire. More information on their professional background can be found in the next section. Some screenshots of this questionnaire can be seen in Figure 13. The full version is available in the Appendix, in section 15.1.

Questionnaire for the media industry (*MediaProfQuest*): This questionnaire was targeted at media industry representatives (news industry, film/TV/radio industry, gaming industry, music industry, content providers, advertisers etc.) as well as at people whose work is about the media sector more broadly (e.g. fact checkers that work in NGOs, representatives of independent authorities that supervise and regulate media, social scientists that study the media, etc.). The survey included the following sections: (1) Professional background of respondents; (2) AI benefits for the media sector; (3) AI risks for the media sector; (4) AI challenges for the media sector; (5) AI strategies & skills in media sector; (6) Adoption of AI solutions by media sector; (7) AI ethics & regulation. In total, 52 people completed this questionnaire. More information on their professional background can be found in the next section. Some screenshots of this questionnaire can be seen in Figure 14. The full version is available in the Appendix, in section 15.2.

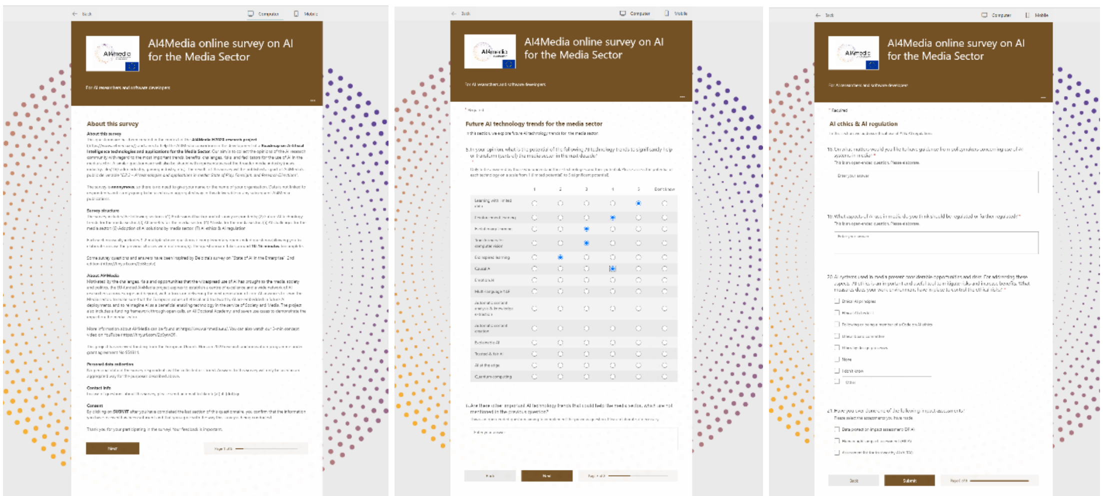


Figure 13: Questionnaire addressed to the AI research community (full version available in the Appendix, section 15.1).

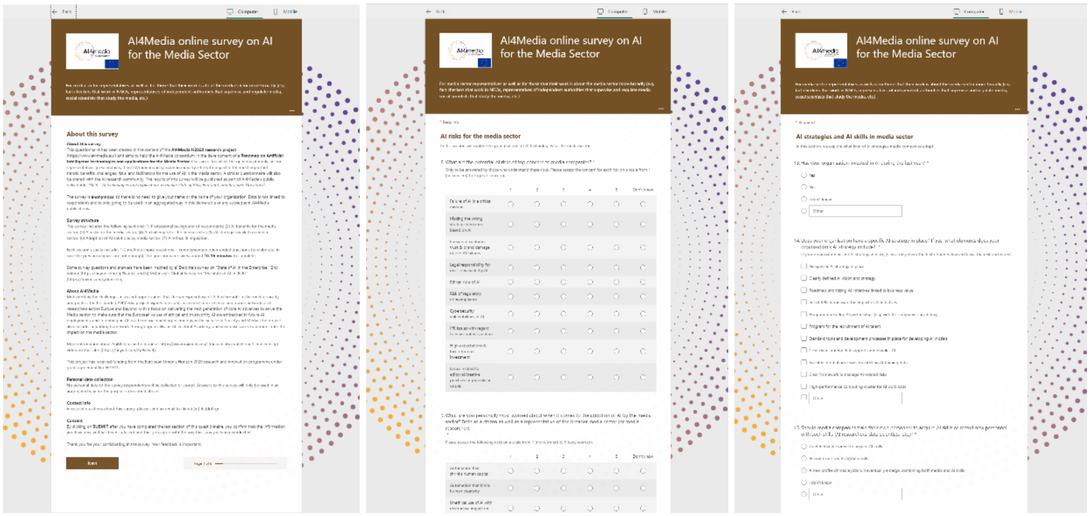


Figure 14: Questionnaire addressed to media industry professionals (full version available in the Appendix, section 15.2).

In both versions of the survey, each questionnaire section usually included 1-2 multiple choice questions + complementary open-ended questions (allowing respondents to elaborate, in case the previous choices were not enough). Some survey questions and answers have been inspired by Deloitte's survey on "State of AI in the Enterprise", 2nd edition[90].

In the following subsections, we analyse the received responses. Each subsection corresponds to a different survey section.

4.1.1 Professional background of survey respondents

With regard to the questionnaire addressed to the ***AI research community***, 81% of the responders work in academia or research and 5% in the ICT industry. 56% are researchers, 10% software developers, and 15% Innovation or R&D managers. Their main research interests include general topics like machine learning (72%), deep learning (65%) and data science (45%), as was expected. With regard to more specific fields, the most popular ones are image/video analysis (55%), text analysis (23%), audio or speech analysis & synthesis (19% and 17%), trustworthy AI (23%) and AI ethics (17%). This information is visualised in Figure 15.

[90] Deloitte's survey on "State of AI in the Enterprise", 2nd edition (2020): https://www2.deloitte.com/content/dam/insights/us/articles/4780_State-of-AI-in-the-enterprise/DI_State-of-AI-in-the-enterprise-2nd-ed.pdf

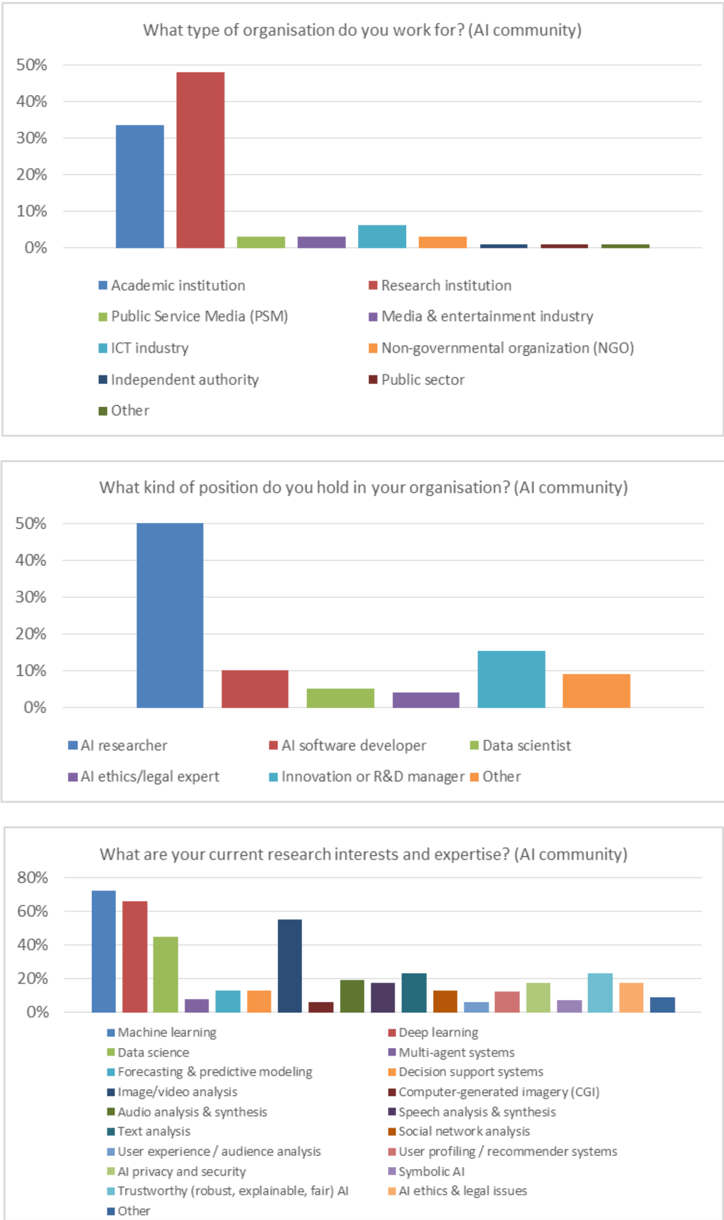


Figure 15: Information about the professional background of respondents to the AIResearchQuest questionnaire.

With regard to the questionnaire addressed to ***media professionals***, 54% of respondents work in Public Service Media (PSM) (25%) or the media & entertainment industry (29%). The rest work for independent authorities, NGOs, academia/research (with their work related to media research), etc. 25% work in a big media network (network that includes several media - TV, radio, news and online media), 19% work in an online news platform while the rest work in various other media organisations (newspaper, TV, Radio, game industry, music, content providers, fact-checking organisation, etc.) or organisations related to media (e.g. PSB associations, regulatory bodies, programmes for data driven media innovation etc.). With regard to their professional background, 21% are journalists, 30% are R&D managers or project managers in a media organisation while the rest are media content creators or providers, game developers, advertisers, fact-checkers, media regulators, media archivers, etc. (see Figure 16).

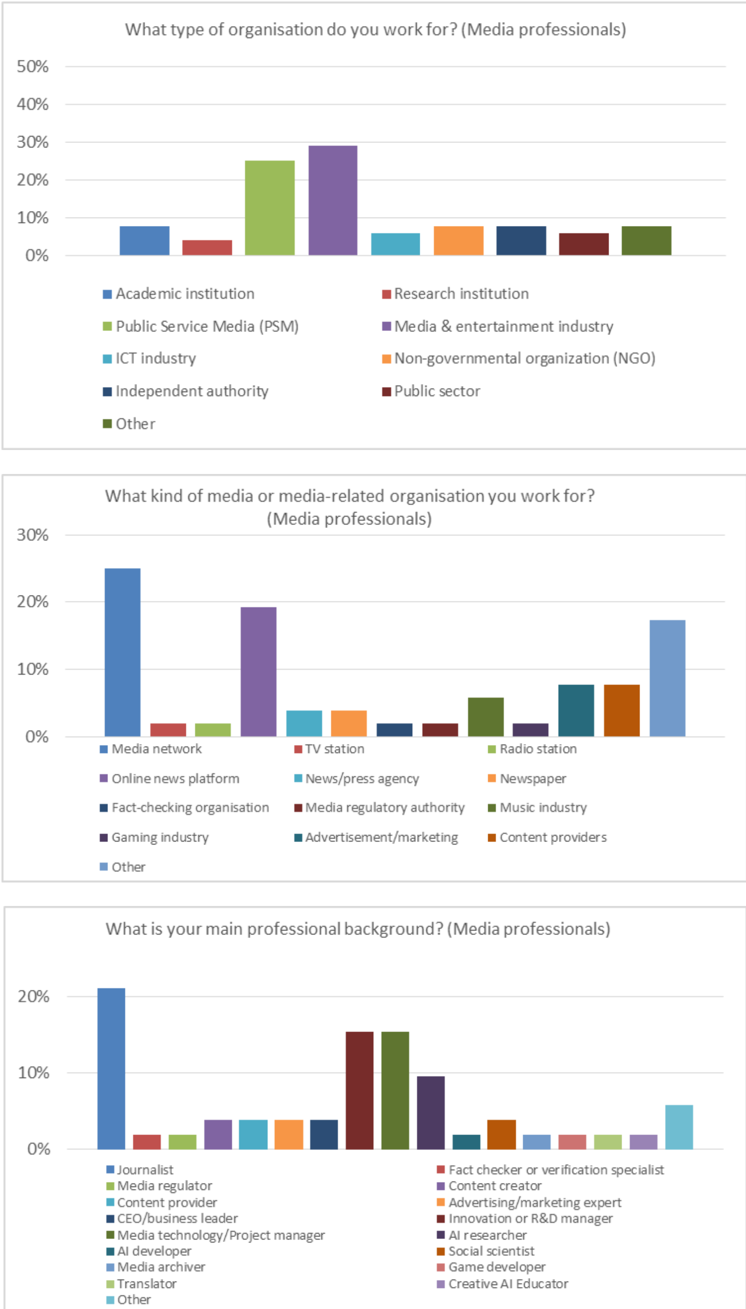


Figure 16: Information about the professional background of respondents to the MediaProfQuest questionnaire.

The 150 respondents worked in 26 different countries in Europe but also in the US and India as can be seen in Figure 17.

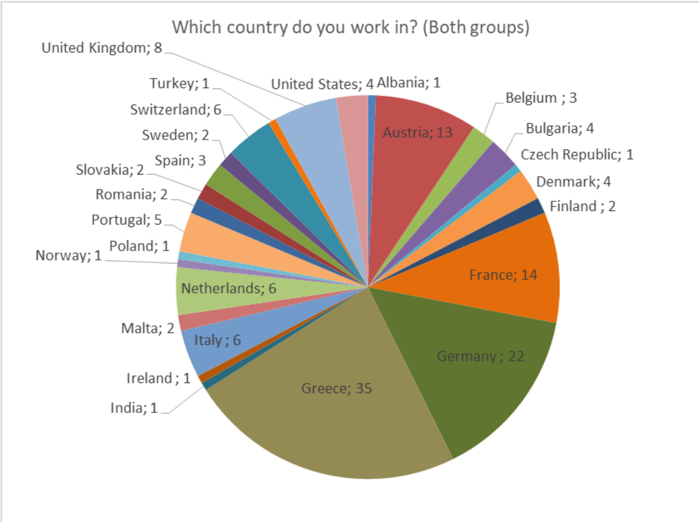


Figure 17: The 26 countries that respondnets of the survey work in.

4.1.2 Future AI technology trends for the Media Industry

The next section of the survey explored the potential of selected **AI technology trends** to significantly help or transform (parts of) the media sector in the next decade. The technologies examined included a selection of the emerging technologies identified in section 3.3 of this report such as reinforcement learning, transformers for computer vision, emotion AI, causal AI, explainable AI, quantum computing, etc. The respondents were asked to assess the potential of each technology on a Likert scale (1: limited potential, 5: significant potential). This question was only addressed to the research community. The results are visualised in Figure 18. We can see that the technologies with the biggest potential according to respondents are automatic content analysis & content creation, multi-lingual NLP, learning with limited data, explainable AI and trusted and fair AI. These answers come as no surprise since these technologies aspire to provide solutions to some of the most pressing problems of the media industry: the ever-growing need for new content, the need to automatically analyse content to extract knowledge, the language barrier, the lack of data to train algorithms for new tasks or new domains, and the need for AI that can trusted to be fair and is able to explain its decisions. The technologies that are believed to have less potential are quantum computing and bioinspired learning, which are at the same time the two technologies for which we got a high number of “I don’t know” answers. Both technologies are still in a very experimental stage and quite far from becoming widely used in the media or other fields, which explains why respondents may not see their potential to change things significantly in the next decade.

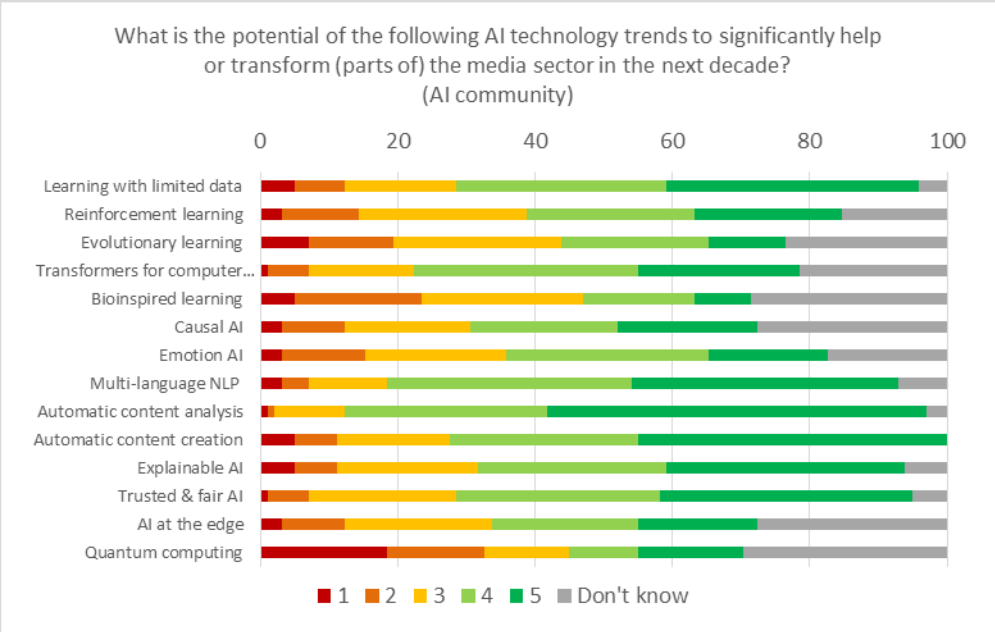


Figure 18: Potential of emerging AI technology trends for the media sector (Likert scale: 1= limited potential, 5 = significant potential).

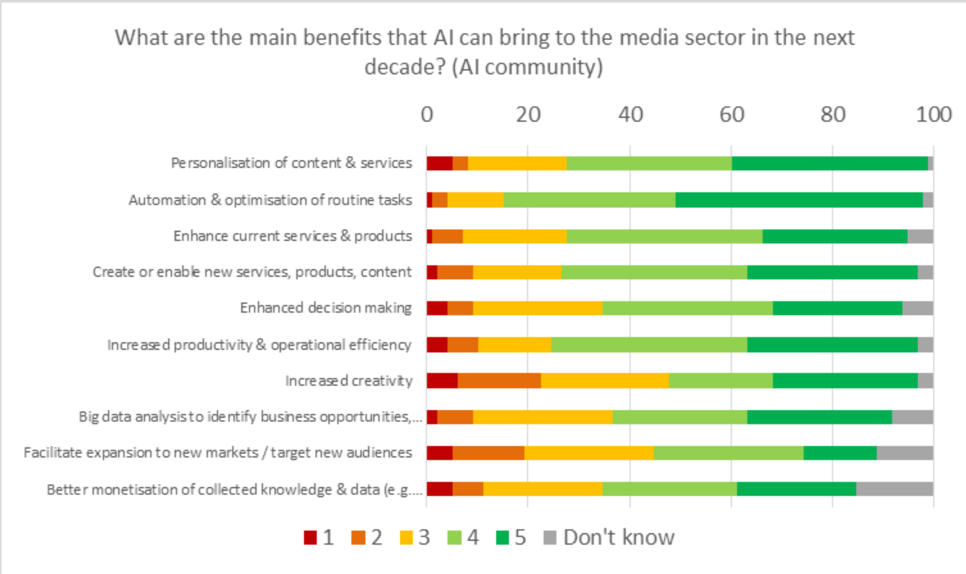
This section also included an open question, focusing on which other emerging technologies have high potential for transforming the media sector. Answers included:

- Transfer learning;
- Federated learning & privacy preserving technologies;
- Generative AI;
- Better knowledge leveraging (particularly hierarchical and contextual) in training data;
- AI exploiting multi-modality (combining text, image, sound, sentiment, context);
- Thinking Models instead of Language Models;
- MLOps to deploy and maintain ML models in production reliably and efficiently;
- Green AI, scalable and energy-efficient technology.

4.1.3 Benefits of AI technology for the Media Industry

This section of the survey aims to explore the potential of AI to positively affect and benefit the media sector. Respondents were asked to assess each **potential benefit** for the media sector on a Likert scale (1: not important, 5: very important). The list of potential benefits that AI can bring to the media sector was devised based on the state-of the-art analysis of section 3 of this report – some answers were also inspired by Deloitte's survey as mentioned above. The focus was on popular applications and general business benefits that are applicable to all the different sectors of the media industry. Thus, we examine how important are from a business perspective personalisation of content & services, creation of new services or content, automation of routine tasks, enhanced decision making, big data analysis to identify business opportunities, improved monetisation of content and data etc.

This question was asked to both the AI and media communities. The results are visualised in Figure 19.



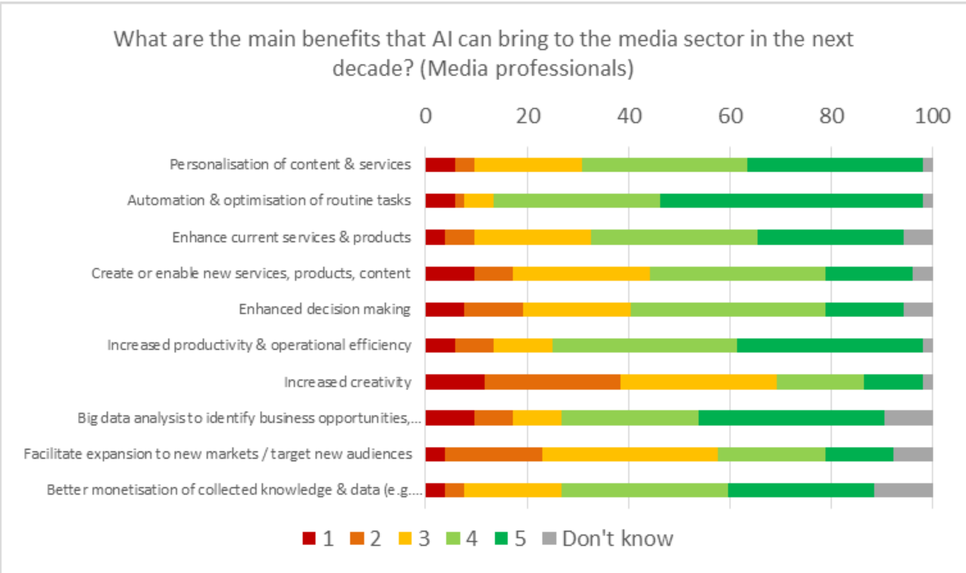


Figure 19: Assessment of the potential AI benefits for the media sector by a) members of the AI research community (top) and b) members of the media industry (bottom) (Likert scale: 1 = not important, 5 = very important).

We can see that survey participants consider automation & optimisation of routine tasks, personalisation of content and services, increased productivity & operational efficiency, and enhancement of current services as the most important benefits of AI. On the other side, both types of respondents seem to believe much less in the potential of AI to increase creativity or facilitate expansion to new markets / target new audiences. The finding about creativity is not that surprising since in other surveys too AI is considered as an assistant - someone to do the hard work - that will increase creativity only indirectly by freeing the workforce from tedious and boring tasks. People seem to be still suspicious of this aspect of AI since creativity (also related to feeling, inspiration, innovative ideas, etc.) is mainly a characteristic attributed to humans. It is interesting to observe that media professionals are twice as sceptic about this issue compared to AI researchers. Most "I don't know" answers have been received with regard to AI's potential for monetising media assets, identifying business opportunities, and facilitating expansion to new markets. Obviously, these are complex issues that involve a lot of different factors (financial, social, business) and many media professionals may not be sure how exactly AI can fit in this equation (note however, that for all of these potential benefits, positive responds are much higher). As a general observation the results show that AI researchers are more eager to believe in AI's transformative role than the media sector is, although both groups undoubtedly acknowledge that it will significantly help the media sector.

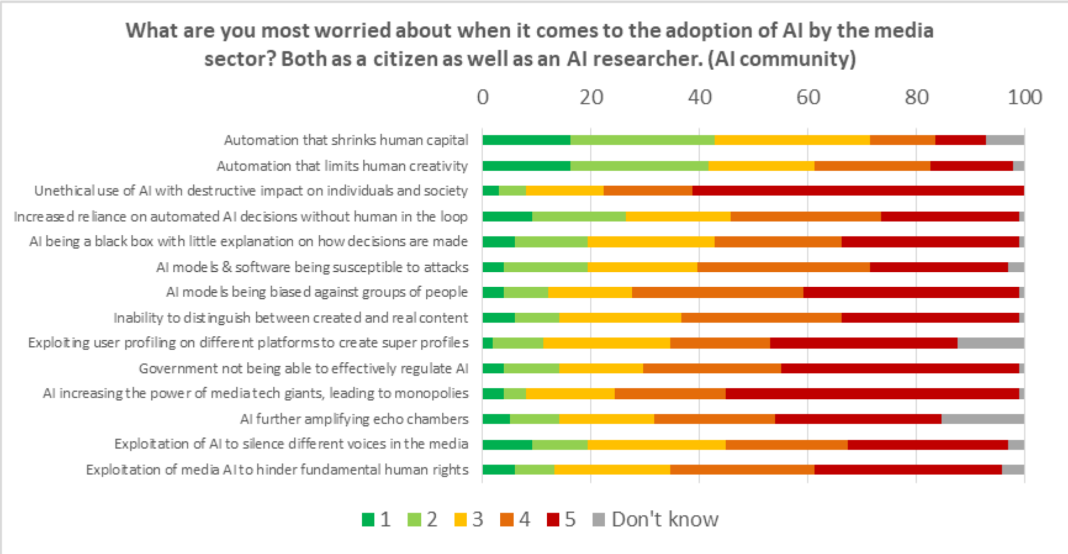
This section also included an open question, focusing on what other benefits AI can bring to the media sector besides those included in the survey. The potential of AI to remove language barriers, automation of tedious tasks and content creation were the topics discussed more by the respondents. Some other interesting answers include:

- Leveraging existing content to repurpose it (learning from videos, for instance) will enable a lot of new products.
- Multilingual and transnational cooperation in content development and journalistic work.
- Democratisation and quality content generation by everyone and everywhere.
- Reducing marketing costs and providing the best possible ROI.
- Solid ground decisions on what content to create and when to release it (content and timing).
- Increase accessibility for disabled people, information accessibility and organisation for users, researchers' research.
- Information fact-checking and content verification, regardless of language.
- Enhanced VR applications for all media.

Finally, media professionals were asked whether they believe that the use of AI has a clear value or benefit for their media organisation. 71% answered 'Yes', 6% 'No', 6% 'I don't know', while 17% said that it was not clear to them. Although, there is definitely a positive expectation for AI, 23% of respondents are still not sure about how AI can or whether AI will benefit their organisation. This necessitates increased awareness campaigns about what AI is and what it can do and possibly more clear strategies for the adoption and operationalisation of AI across a media organisation.

4.1.4 Risks of AI technology for the Media Industry

This section of the survey aims to explore the **potential risks of AI technologies** for the media sector. Respondents were asked to assess each potential risk on a Likert scale (1: not worried, 5: very worried) both as professionals in their field but also as citizens. The list of potential risks that AI can bring to the media sector was devised based on the state-of the-art analysis of section 3 and section 11.1 of this report. We examine risks of AI automation, unethical use, lack of explainability, increased reliance on AI decisions, bias, vulnerability to attacks, user profiling etc. This question was asked to both the AI and media communities. The results are visualised in Figure 20.



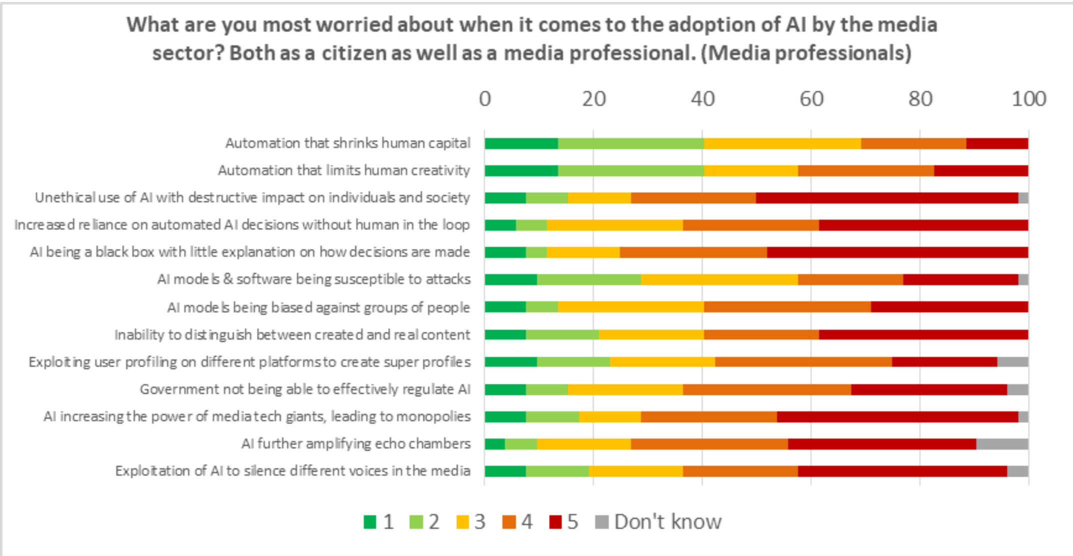


Figure 20: Assessment of potential AI risks for the media sector by a) members of the AI research community (top) and b) members of the media industry (bottom) (Likert scale: 1 = not worried, 5 = very worried).

We can see that survey participants worry less about the risks of automation with regard to the loss of human jobs or limitation of human creativity. But they seem to worry a lot about everything else, most importantly about the unethical use of AI and its impact on society, AI bias, AI's lack of explainability, and potential exploitation of AI to hinder fundamental human rights. It is interesting to note that media professional consider the former (unexplainable AI) a bigger risk than unethical use of AI. Media professionals also worry more about AI amplifying echo chambers, being used to silence voices in the media as well as about increased reliance on AI decisions - while researchers worry that governments will not be able to effectively regulate AI, which will of course affect AI research as well. We also spot three areas where there is a non-negligible difference between the opinions of media professionals and AI researchers. Media professionals worry much more about reliance on AI decisions and non-explainable AI and much less for AI models & software being susceptible to attacks than AI researchers do. This is not hard to explain. AI researchers develop AI so they know more about how it works or how it can be attacked. Media professionals don't and that is why they worry about the increasing reliance on AI.

In addition to the previous question, media professionals where asked about the potential AI risks of top concern to media companies. The answer options were mostly inspired by the Deloitte survey. Note that here the focus is on the business side (what are the risks for the company) while the previous question also included a societal dimension, asking respondents to provide their opinion also as thinking citizens. The responds to this question are summarised in Figure 21. Risks of top concern include high expectations & low return, failure of AI in a critical mission, and ethical risks. Especially the first, about the hype of AI and what it can actually deliver, seems to be the most important issue: making big investments on AI having high expectations but then not receiving a lot of benefit in return. Failure in a critical mission is also important since one of the main motives for AI adoption is the potential for automation of various tasks – this can be both a blessing and a curse. For the news industry, risks related to editorial/creative practices and journalistic values are also high in the list of concerns. Media professionals seem to worry less about making the wrong strategic decisions based on AI, non-compliance with regulations or IPR issues related to new AI-generated content. Since adoption of AI in the industry is still in early stages, such issues may not worry people yet, especially since they still are not clear about the extent or repercussions of these risks (please note that for some of these risks we have received the most "I don't know" answers).

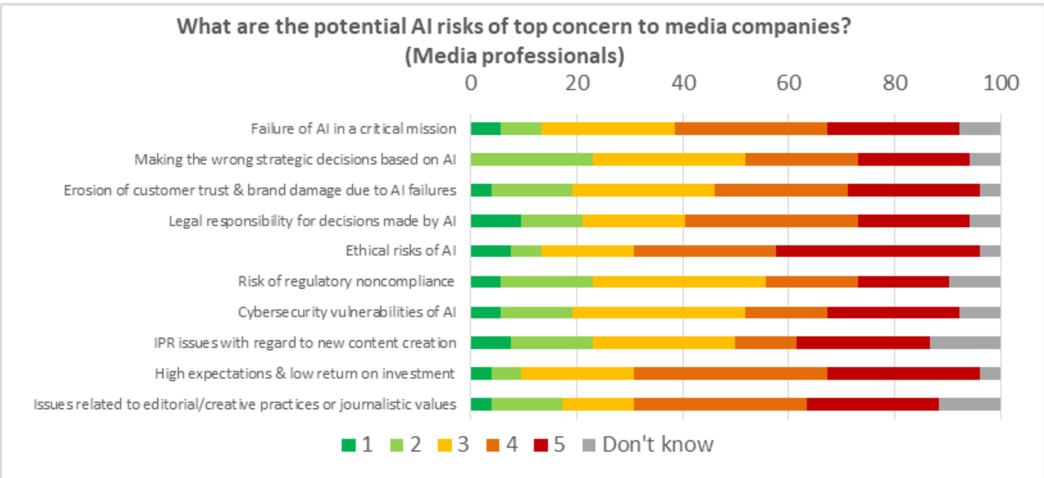


Figure 21: Assessment of top AI risks for media companies (Likert scale: 1 = no concern, 5 = great concern).

Finally, this section also included an open question, asking respondents to identify other risks not mentioned in the previous questions, elaborating on issues related to their own research area or media organisation. Creativity, diversity, bias, employment, access, and IP are among the issues mostly discussed by respondents. Some interesting answers include:

- Devaluing of human creativity due to mass production enabled by AI and reducing cultural diversity.
- More societal gaps (technology, digital competences, access to content and information, etc.) introduced.
- People with no AI skills or unable to adapt may see their salary decreasing.
- AI that misbehaves or AI that cannot be trusted (which can erode trust in media).
- We should understand the full impact of bias in AI before adopting it. Humans tend to hide personal responsibility behind systems.
- Wrong AI-based decisions are a serious threat, even small decisions may be biased.
- Massive IP infringement. Unauthorised use of IP as training material, or input material. Untraceable unlicensed use. Unlicensed derivative works created by AI.

4.1.5 AI challenges for the Media Industry

This section of the survey aims to explore the ***challenges encountered by AI researchers/developers*** when developing AI technologies for the media sector and the challenges encountered by media professionals with regard to the adoption AI. A different list of challenges was assessed by each group of respondents since the former develop AI while the latter are the users of this technology. The lists were devised based on the state of the art analysis of sections 3 and 11 of this report as well as by discussions that have been happening in the consortium since the beginning of the project with regard to the challenges faced by both research/technical partners and media industry partners. Respondents were asked to assess the importance of each challenge on a Likert scale (1: not important, 5: very important). The results are visualised in Figure 22 and Figure 23, respectively.

With regard to the AI research community, we examine challenges such as lack of data, lack of understanding media sector needs, compliance with regulations, ethical challenges, lack of funding and talent, big-tech monopoly, etc. In Figure 22, we can see that AI researchers and developers consider by far the greatest challenges the lack of data in general to train and test their algorithms but also the reluctance of the media industry to share their data to help and accelerate AI research. The US-China big tech monopoly in AI software and data, facilitated by enormous funding and generally much less regulation, is also something that creates challenges especially for European researchers. Other important challenges include the lack of open access culture for sharing AI algorithms, lack of effective benchmarks and to a lesser extent lack of funding to develop AI for media. What is least challenging or worrisome is apparently the lack of AI talent in universities, research centres and ICT industry or collaboration with media industry and understanding of its needs.

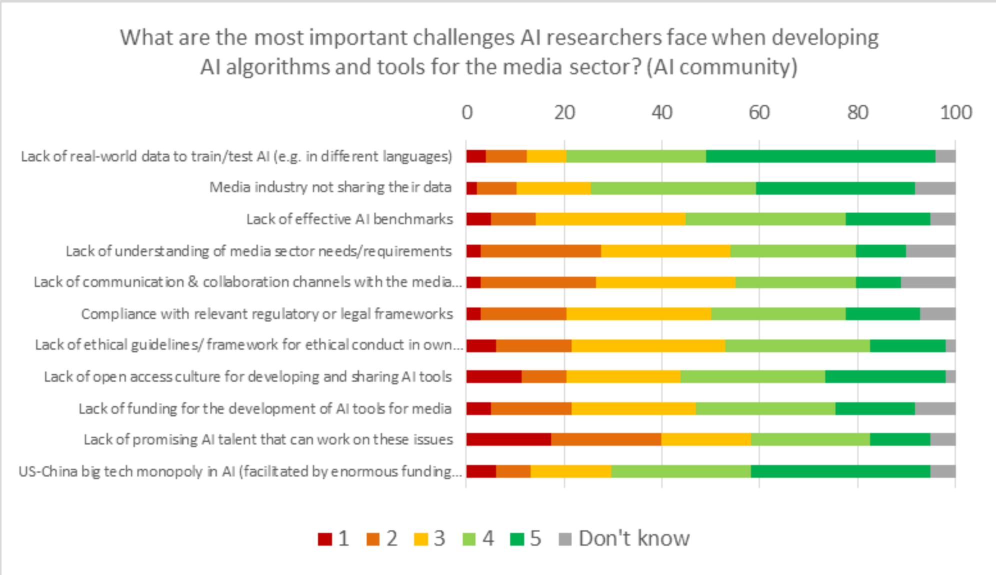


Figure 22: Assessment of most important challenges researchers face when developing AI algorithms and tools for the media industry (Likert scale: 1 = not important, 5 = very important).

This section also included an open question, asking respondents to describe challenges not mentioned in the previous question, elaborating on issues related to their own research area. Access to compute infrastructure, a more multi-disciplinary approach to AI development, and the public research sector are among the issues discussed by respondents. Some interesting answers are listed below:

- Access to computing resources.
- Getting a realistic representation of AI capability embedded in the media.
- Lack of reliability of open source tools as well as real-world data benchmarks.
- Consideration of full AI life cycle (incl. testing, certification, monitoring, etc.) in R&D ecosystem.
- Lack of involvement of media professionals and social scientists in AI development.
- Lack of proper education and sensitisation on non-technical issues (i.e., ethics, aesthetics, human behaviour, etc.).
- Limited usage of media archives because high-level executives in both worlds do not care about reciprocal existence (AI for archives - archives for AI).
- Lack of attractiveness of public research sector for AI research talent vs. non-EU corporations.

With regard to **challenges encountered by the media community**, on the other hand, we examine challenges such as understanding what AI has to offer, operationalising AI and measuring its business value, cost of developing AI solutions, AI explainability, ensuring ethical use or compliance with regulations, lack of AI skills in the media industry, etc. In Figure 23, we can see that media professionals consider by far the greatest challenges the integration of AI in business operations and processes, understanding what AI can do to improve or facilitate workflows and processes in the media industry, and lack of relevant skills in media professionals. The US-China big tech monopoly is also a challenge, since these companies seem to want to suck the oxygen out of the traditional media industry by becoming the main providers of online content and thus increasing their profits and users. Other important challenges include AI explainability as well as data access and privacy. The least important challenges, according to media professionals, are the establishment of an internal ethical framework for AI use, attracting AI talent in the industry, or proving/measuring AI's business value.

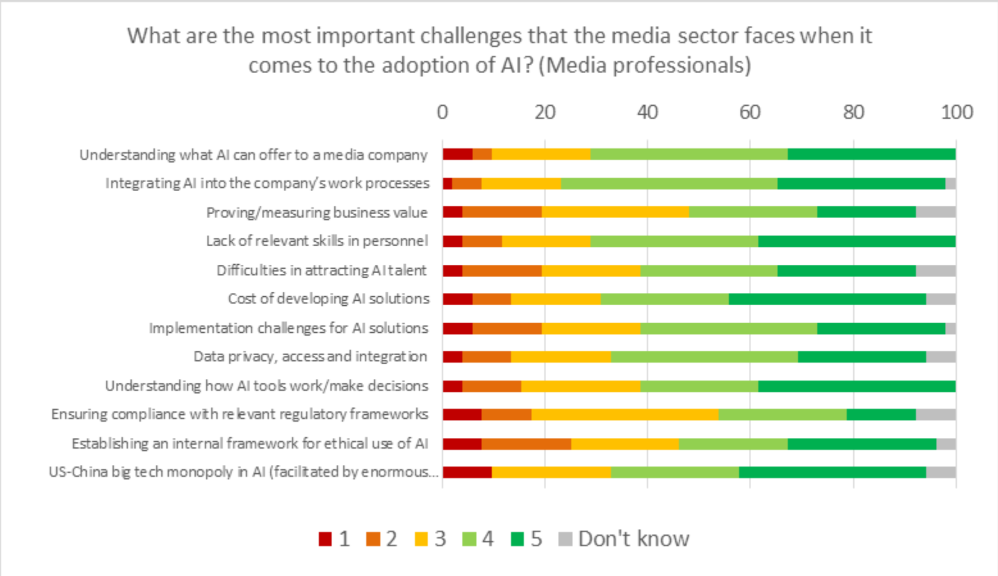


Figure 23: Assessment of most important challenges the media industry faces when when it comes to the adoption of AI (Likert scale: 1 = not important, 5 = very important).

The same open question was addressed to media professionals, asking respondents to describe challenges not mentioned in the previous question, elaborating on issues related to their own line of work. Compliance with upcoming EU regulations, need for transparency, IP issues, or how AI is currently integrated in a media organisation are some of the issues discussed by the respondents. Some very interesting answers are listed below:

- Establishing an internal framework for ethical use of AI is the 'easy' part. It is much more difficult to develop, purchase, integrate and maintain AI systems/tools within the media company (both in-house and from third-parties) that live up to these (paper-based) frameworks in terms of actual implementation and end user acceptance - due to the comparatively early and still immature status of transparent and trustworthy AI, and due to new regulation frameworks coming up.
- Need for content to have transparency and clear labels.
- Getting technology companies to respect copyright, rather than claiming training and other uses is "fair use" or similar.
- News is an ever changing landscape and therefore frequent update of AI models may be needed.

4.1.6 AI strategies and skills in the Media industry

This section of the survey includes four questions that aim to explore what kind of AI strategies are adopted by the media industry and also examine the issue of AI talent. The questions have been addressed to media sector professionals (except one that is addressed to both targeted groups of responders).

The first question was about **investments on AI**. Respondents were asked whether their organisation has invested in AI during the last year. Only 50% of responders answered positively. One third (33%) said that their organisation had not invested in AI while 9% said they didn't know (Figure 24). One responder answered that they have in-kind investment through participation in projects and testing (a news agency), while another one said that an investment is planned. The respondents that answered "No" were mainly affiliated with NGOs and independent authorities (e.g. regulating media). Among those working in PSM or the media and entertainment industry, 73% have invested in AI.

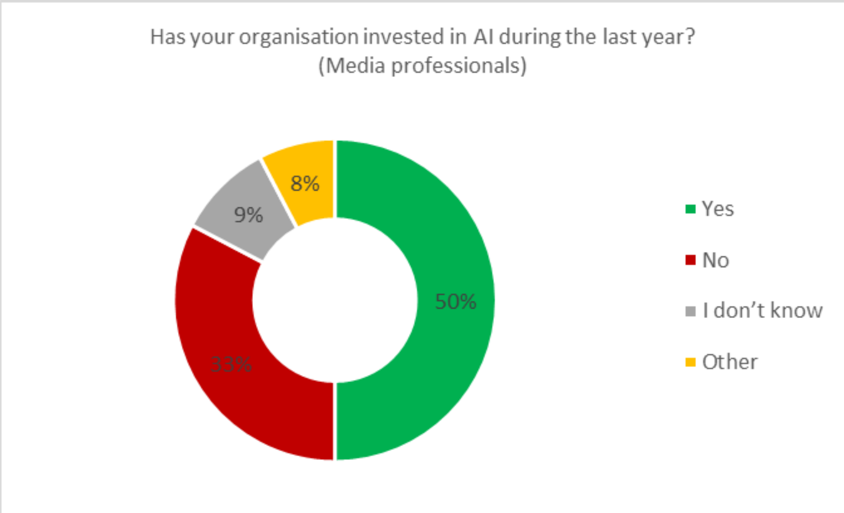
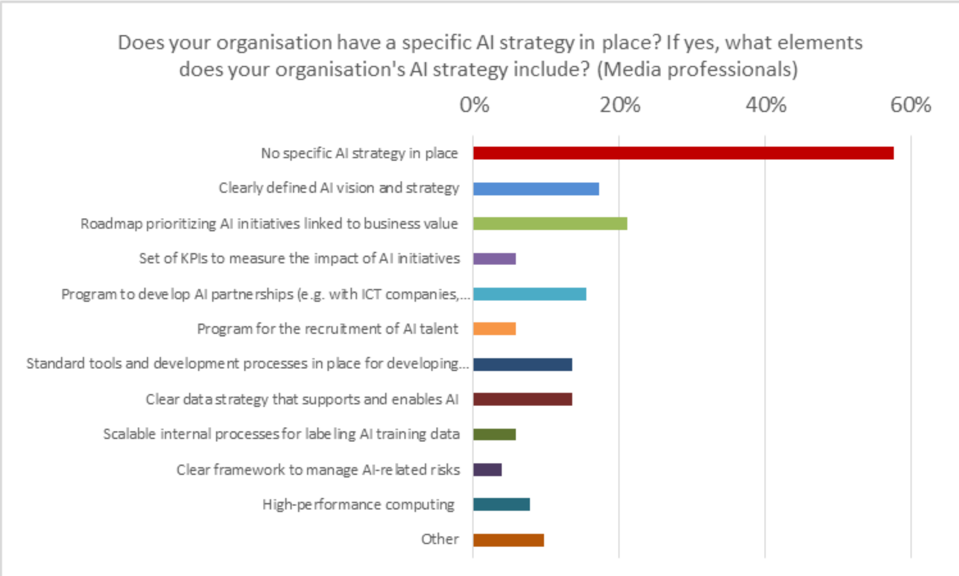


Figure 24: Investement in AI by media and media-related organisations.

The second question examined what kind of **AI strategies** media and media-related organisations have in place, if any. The responders were provided with a list of options from which they could select more than one. The available options include, for example: clearly defined AI vision and strategy, roadmap prioritising AI initiatives linked to business value, KPIs to measure the impact of AI initiatives, program to develop AI partnerships (e.g. with ICT companies, academia), program for the recruitment of AI talent, standard tools and development processes in place for developing AI models, clear data strategy that supports and enables AI, scalable internal processes for labelling AI training data, clear framework to manage AI-related risks, and high-performance computing cluster for AI workloads. This list has been inspired by McKinsey's Global Survey on "The state of AI in 2020"[91]. Results are visualised in Figure 25. An astounding 58% of respondents answered that their organisation had no specific AI strategy in place (see Figure 25(a)). Only 42% had some kind of AI strategy.



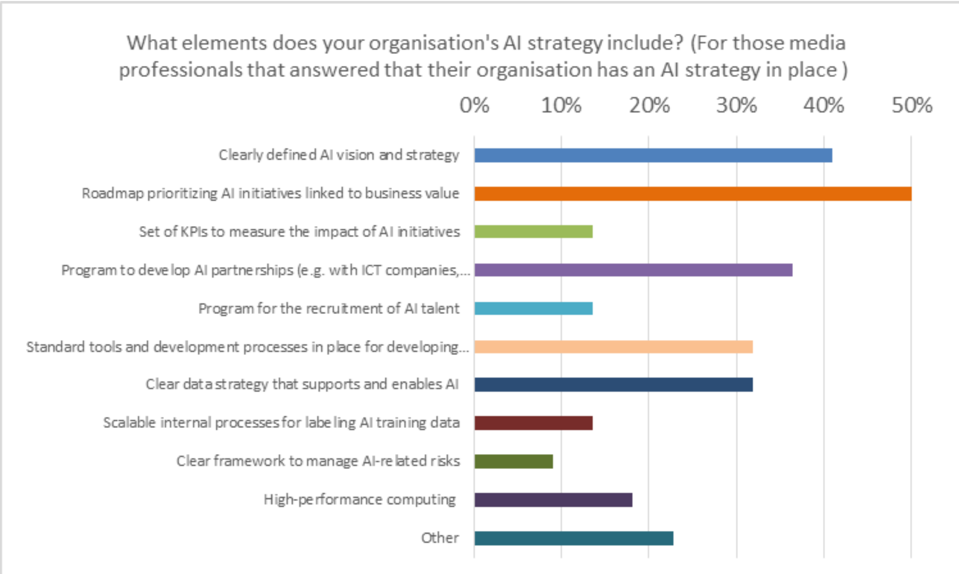


Figure 25: AI strategy of media and media-related organisations: a) all answers are shown, including "No AI strategy in place" (top); b) Those that have answered "No AI strategy in place" have been excluded from the analysis – the percentages refer only to those organisations who have some kind of AI strategy (bottom).

[91] McKinsey's Global Survey on "The state of AI in 2020" (2020): https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Analytics/Our%20Insights/Global%20survey%20The%20state%20of%20AI%20in%202020/Global-survey-The-state-of-AI-in-2020.pdf

Trying to examine what kind of elements the adopted AI strategies have, we find that out of those who have reported some type of AI strategy (the aforementioned 42% of total respondents), 50% have a roadmap prioritising AI initiatives linked to business value, 41% have a clearly defined AI vision/strategy, 36% have a program to develop AI partnerships, 32% have standard tools and development processes in place for developing AI models, and 32% have a clear data strategy that supports and enables AI. The most amazing finding is that only 9% of those with an AI strategy in their organisation have a clear framework to manage AI-related risks (incl. ethical risks). That is approx. 4% of the total respondents. This clearly poses a significant danger for the organisation and its employees and has the potential to damage public trust in media, trust that is continuously eroding also affected by other factors. Similarly, only 13% of those with an AI strategy answered that their strategy includes specific KPIs to measure impact of AI in organisation and only 13% have internal processes for labelling AI training data. Not being able to measure the actual impact and added value of the AI tools in the organisation may hinder further adoption of the technology in the long run (why pay for something if you do not know whether it actually works). Also, lack of processes for data annotation contributes to the problem of data availability mentioned already in the previous sections. With regard to the type of organisation, 46% of PSM and 46% of media and entertainment industry companies, 62% of large media networks and 30% of online news platforms have an AI strategy.

The next question was about whether media companies should ***train their own personnel to acquire AI skills or recruit new personnel with such skills*** (AI researchers, data scientists, etc.). As shown in Figure 26, opinions are split on this issue: 29% believe media organisations should recruit experts with AI/data skills, 27% believe they should train existing personnel to acquire AI skills, and 27% think that a new profile of media jobs will eventually emerge, combining both media and AI skills. Among those that provided alternative answers some said that a combination of the above would do the work, while one respondent pointed out that it all comes down to the budget available for this.

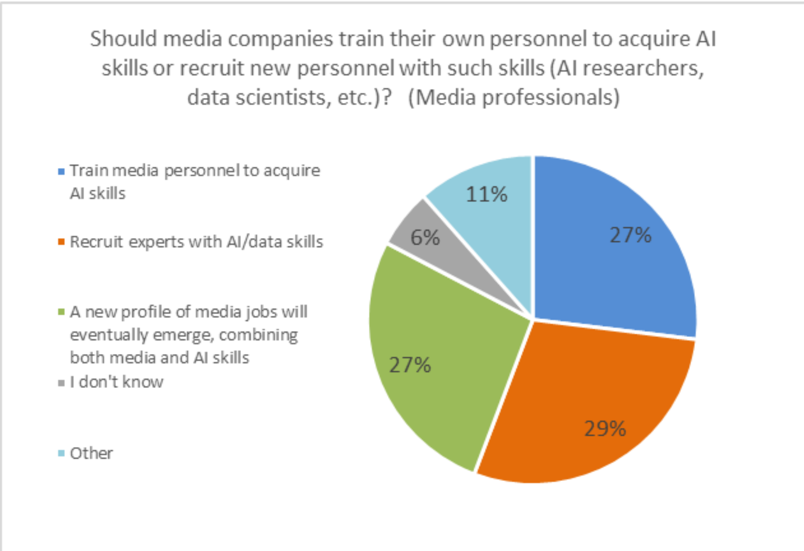


Figure 26: How to increase AI skills in the media sector.

Finally, the last question, addressed to both media professionals but also the AI community, complemented the previous one, examining **what kind of skilled personnel a media company needs to hire** to overcome the often encountered AI skills gap. The results are visualised in Figure 27. AI software developers (69%) and data scientists (67%) were by far the most popular choices among media professionals, followed by experts on AI ethics and legal issues (44%). For AI researchers, the most popular choice is AI software developers (64%), followed closely by data scientists (53%), AI researchers (52%), domain experts (51%) and ethics experts (50%). The most important difference between the two groups concerns domain experts. 33% of media professionals consider that they are necessary in contrast to the 51% of the AI community. Since media professionals are in many cases also domain experts they do not see the need to hire more such experts. However, in the research community domain knowledge is a big need, with domain experts usually in high-demand. This survey result seems to reflect this reality. Other types of personnel not included in the list but mentioned by respondent include representatives of democratic society (e.g. NGOs) but also experts that are able to effectively integrate AI into media workflows and people that know how to develop energy-efficient software tools.

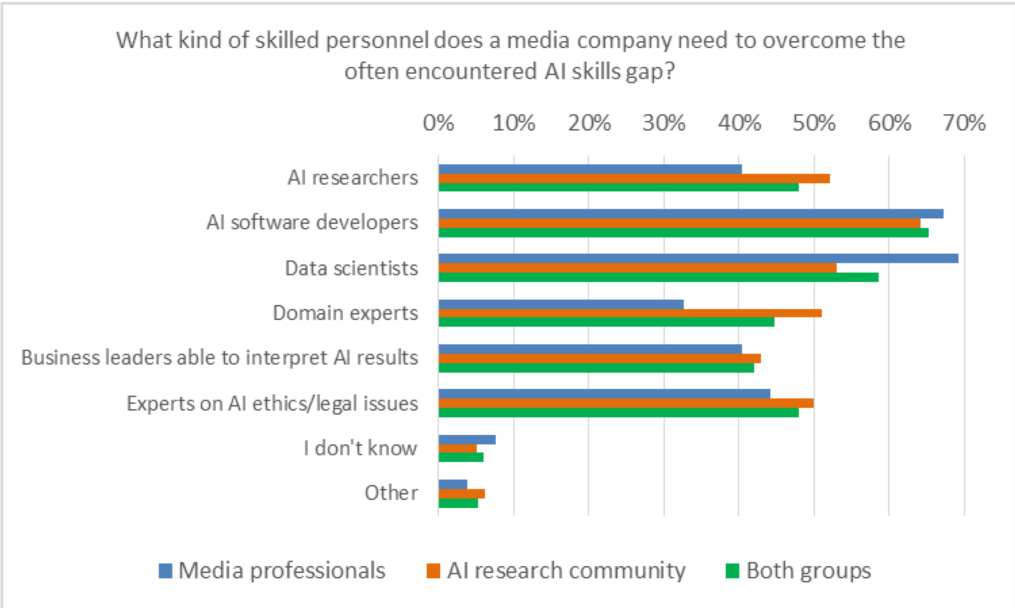


Figure 27: Skilled personnel that media companies need to hire to overcome the often encountered AI skills gap. Answers provided by responders from the media industry, the AI research community and both.

4.1.7 Adoption of AI solutions by the Media Industry

In this section of the survey, we explore adoption of specific AI technology applications by the media sector as well as the requirements that may facilitate further adoption of AI technologies.

The first question examines what kind of **AI/ML-enabled applications** most media professionals use. As can be seen in Figure 28, the most popular applications are image/video analysis and NLP (44% each), followed by audio analysis (35%), user experience/audience analysis (31%).

Social media network analysis (27%) and recommenders (25%) are also used by many professionals. The use of fact checking tools is also considerable but that can be attributed to the fact that 21% of respondents are journalists. Applications with low penetration are automatic-decision making for business (6%) and market analysis/forecasting (10%). The former can be attributed to the fact that media professionals still do not trust AI to make decisions without a human in the loop and also that AI is not mature enough to be widely used for such applications. The latter is also probably attributed to this but also to the fact that our pool of respondents does not include a lot of representatives from media sectors where such technologies are more mainstream, like advertising, music/film, big streaming platforms or social networks. Note also how the use of automatic content creation tools, although considered one of the most promising fields of AI research, is still currently limited. Other technologies (not in the list) mentioned have mainly to do with AI for security applications.

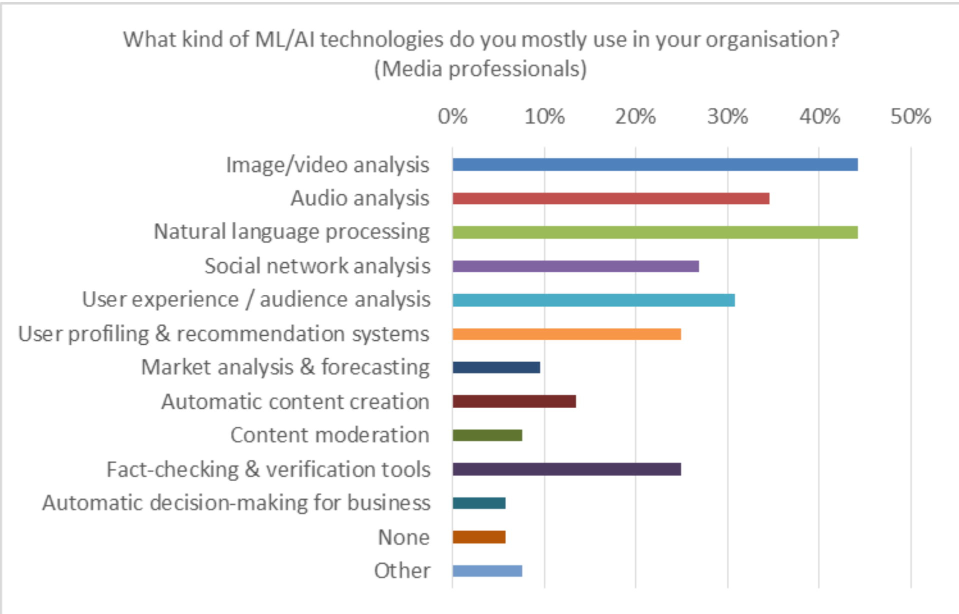
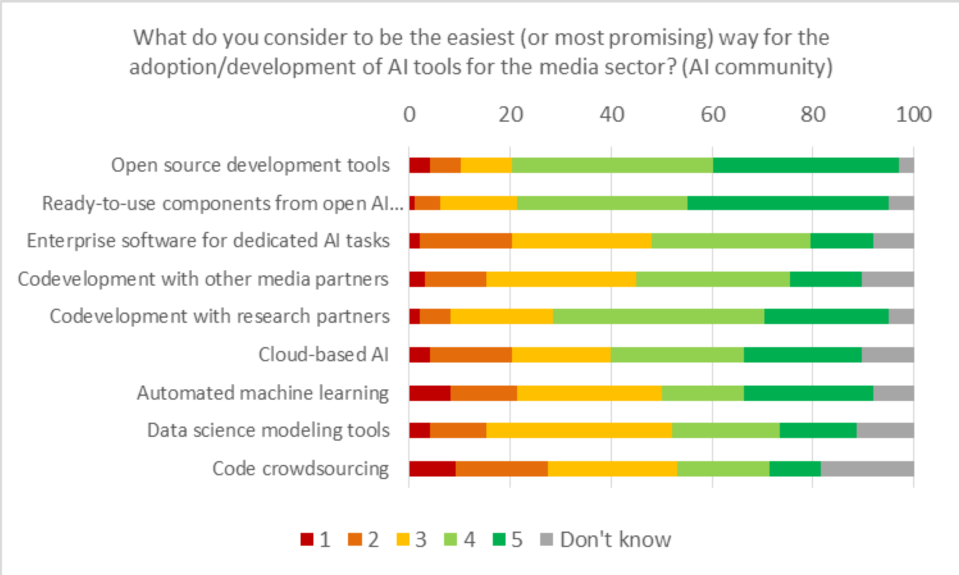


Figure 28: AI/ML-enabled applications used by media professionals.

The next question explores the **most promising ways for the easy adoption of AI** by the media sector. The list of answers to this question includes solutions like open source development tools, ready-to-use components from open AI repositories, enterprise software for dedicated AI tasks, co-development with other media or research partners, cloud-based AI, automated machine learning and data science modelling tools, and code crowdsourcing. The question and answers have been inspired by a similar question in the Deloitte survey. The respondents were asked to assess how easy it is to use each solution on a Likert scale (1: difficult, 5: very easy). The question was addressed to both AI researchers and media professionals and the results are visualised in Figure 29.



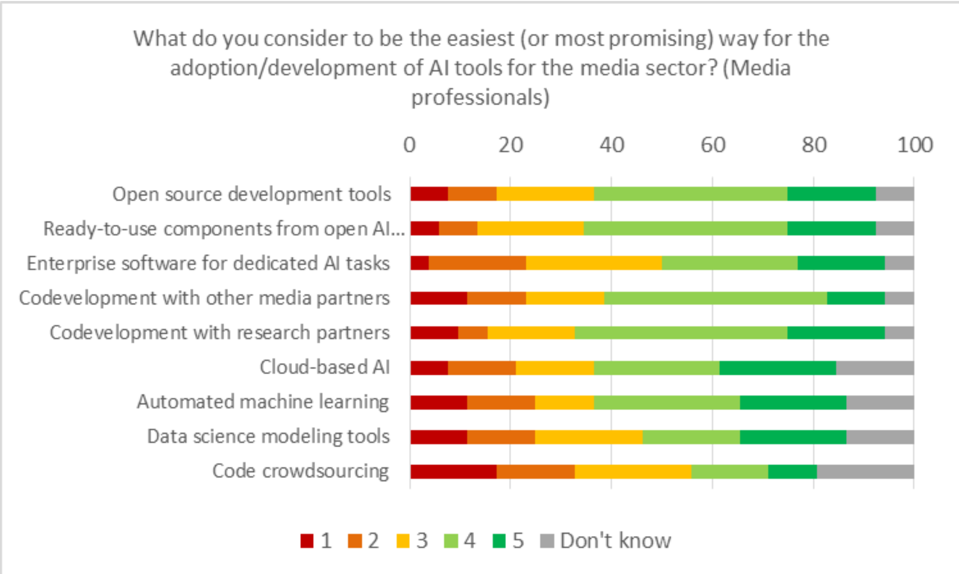


Figure 29: Most promising solutions for the easy adoption/development of AI tools in the media assessed by a) members of the AI research community (top) and b) members of the media industry sector (bottom) (Likert scale: 1=difficult, 5=very easy).

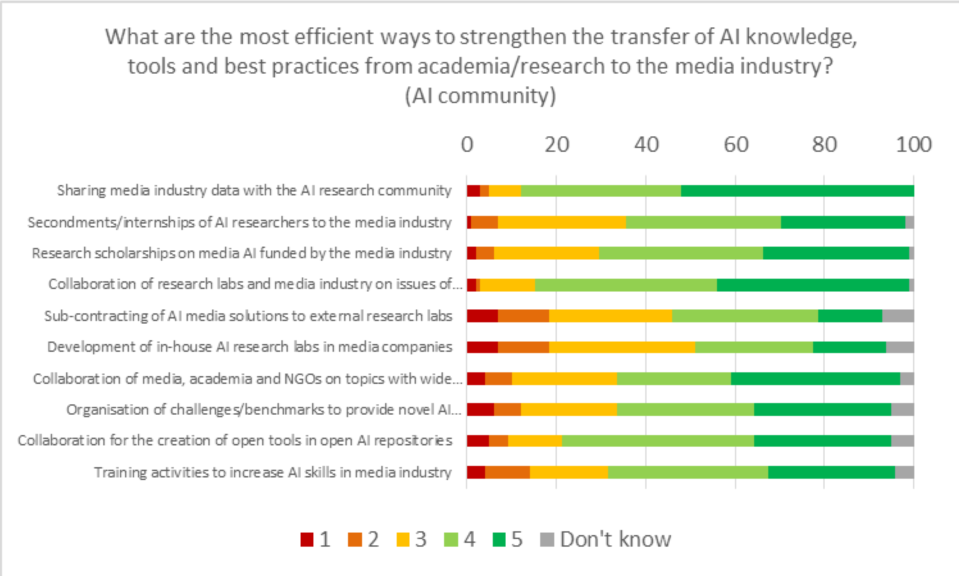
Both groups of respondents consider open source development tools, ready to use components from AI repositories and co-development of media with research partners to be the most promising solutions. The first two are by far the most popular among AI researchers while the third is the most preferable solution among media professionals, closely followed by the other two but also by co-development with other media partners.

The least favourable solution is code crowdsourcing, followed by enterprise software, automated machine learning for media professionals, and at some extent cloud-based AI. Code crowdsourcing also received by far the most “I don’t know answers”. Such approaches although desirable in order to increase the democratisation of AI development have inherent difficulties and have yet to prove that they can offer efficient and trustworthy solutions applicable in the industry. The answers of the respondents show a clear preference for open solutions for both AI development and deployment and for the co-development of solutions based on cooperation between researchers and industry or on collaborating clusters of media organisations. This is also confirmed by the way respondents evaluated enterprise software and automated tools, for which one cannot really know how they work or how they were trained or if they can be trusted since they are usually black-boxes performing a specific task. The need for AI that you can trust and for AI that you know how it has been developed is confirmed by the analysis of responses.

The previous question was also complemented by an open question, asking respondents to describe additional solutions not mentioned in the previous question, also taking into account their own experiences. Respondents insisted on the need for media-technology partnerships for AI development, also including potential users; increased leadership by broadcast or other media associations like EBU to guide the development of new AI tools; but also on scalable technologies and tools that could facilitate link existing media workflows with AI software. Some interesting views are listed below:

- Co-development with technology providers. The cost to build/train large scale models like GPT3 is outside the realm of possibility for most media companies.
- Co-design incl. participation of potential users, highest quality-of-service, proof of added-value, user-friendly navigation.
- Inclusion of highly interdisciplinary approaches, regarding epistemology, cybernetics, neuroscience, psycholinguistics, mathematics, philosophy (new realism) and linguistics.
- Use of AI systems developed by/within Industry Associations for their members.
- The EBU should be a "guiding" organisation that helps understanding, training and offering tools for AI.
- Having events that bring together the media sector and AI researchers. Usually, local versions of these can enable more broad impact, but international events breed better long-term collaborations.
- Making scalable technologies that work through older and newer platforms, energy-efficient tools that can be also applied on 8-, 16- and 32-bit environments.
- Tools for linking existing production and storage workflows to AI software.

The final question of this section of the survey explores the **most efficient ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry**. The list of answers to this question includes solutions like sharing media industry data; secondments/internships of AI researchers to the media industry; research scholarships on media AI funded by the media industry; collaboration of research labs and media industry on issues of common interest; sub-contracting of AI media solutions to external research labs; development of in-house AI research labs in media companies; collaboration of media, academia and NGOs on topics with wide media or societal impact (e.g. AI for disinformation detection); organisation of challenges/benchmarks to provide novel AI solutions; collaboration for the creation of open tools in open AI repositories; and training activities to increase AI skills in media industry. The respondents were asked to assess the potential of each solution on a Likert scale (1: not efficient, 5: very promising). The question was addressed to both AI researchers and media professionals and the results are visualised in Figure 30.



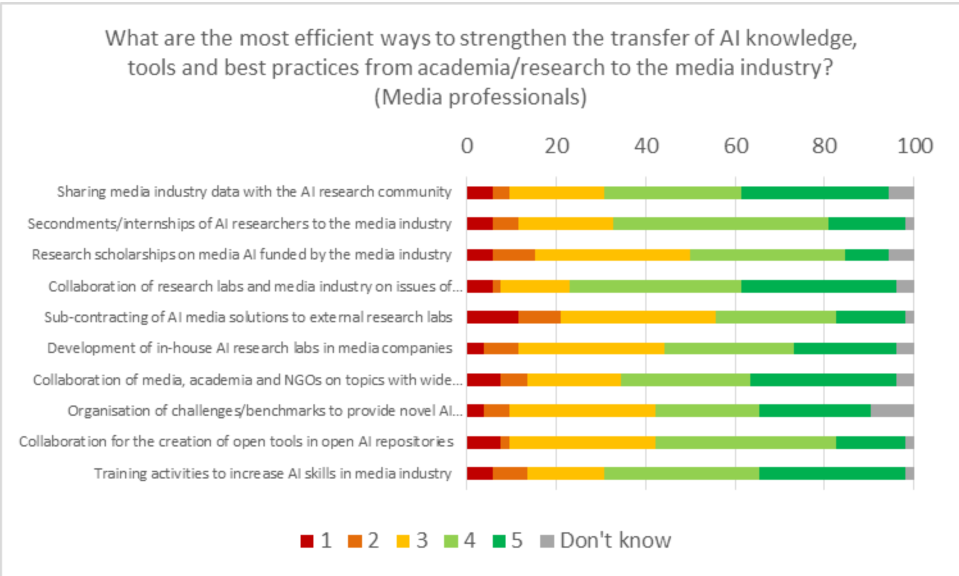


Figure 30: Most efficient ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry assessed by a) members of the AI research community (top) and b) members of the media industry sector (bottom) (Likert scale: 1=not efficient, 5=very promising).

As can be seen from this figure, respondents are generally very positive about most of the proposed approaches. However, some are more favoured than others. For example, by far and large AI researchers believe that transfer of knowledge, tools and best practices can be more efficiently achieved through sharing of media industry data and collaboration between research labs and media industry on issues of common interest, followed by research scholarships on media AI funded by the media industry and collaboration for the creation of open tools in open AI repositories. Media professionals, on the other hand, favour most collaborations of media industry with the research community and training to increase AI skills of media practitioners, followed by secondments/internships of AI researchers to the media industry and sharing of media data. Collaboration between the research and media communities is obviously the most promising approach whatever form it takes. At the same time researchers put special emphasis on data sharing, which as we discussed before is a large impediment in AI development, while media professionals point out the importance of training initiatives so that they are able to follow up and contribute to AI developments instead of being mere bystanders to AI breakthroughs and simple users of the technology others develop. The solution respondents where less enthusiastic about is sub-contracting of AI media tools or applications to external research labs. AI researchers, also do not seem to favour so much development of in-house AI research labs in media companies since that would probably create counter-motives for the collaboration between the two communities.

The previous question was also complemented by an open question, asking respondents to offer additional ideas (not mentioned in the previous list) about the transfer of AI knowledge, tools and best practices from academia/research to the media industry. The main theme of the answers was the need for AI training on all levels (media, academia, society). Some respondents especially pointed out the resistance of parts of the media industry to change or to adopt new technologies. The views and suggestions presented below nicely make the case:

- Training activities to increase knowledge about and acceptance of AI, including aspects of AI trustworthiness.
- Training should start from a very early age; we should rethink and update teaching curricula at University Level, so that soft sciences (Humanities, Social Sciences, Journalism) can catch up on new trends.
- General training for the management level of businesses on the reality of AI currently would be valuable. Several senior leaders in industry are either entirely dismissive or unaware of the current and potential impact that AI/ML is likely to bring on their sector.
- A big obstacle for AI in the media industry is that the industry itself is very traditional and it is hard to change the mentality of decision makers (who are not AI savvy). Unless this changes at the decision-making level, any amount of change from the ground up (tool level and team level) is not going to have a massive impact.
- Exchange of AI researchers between different European countries for a limited time in order to enhance the AI research of countries under development (e.g. under a program like Erasmus for students).
- Permeability of membrane between public/private media sector and academia (à la triple helix approach) with strengthening of public media in the interest of citizens.

4.1.8 AI ethics and regulation

In this section of the survey, we explore AI & ethics as well as AI regulation. The section, addressed to both groups of respondents, includes four questions that aim to shed light on ethics-related practices and processes adopted by industry and researchers, to understand what are the main needs of AI developers and users with regard to guidance from policy makers on the use of AI for the media industry, and to explore the opinions of both communities with regard to AI regulation, especially given the increased legislative/policy making activity of the European Commission on the topic during the last few years (more on this issue in section 11 of this report).

The first question is an open question that aims to examine what matters respondents would like to **have guidance on from policymakers**, concerning use of AI systems in media ("*On what matters would you like to have guidance from policymakers concerning use of AI systems in media?*"). From an analysis of provided answers, the following main areas were identified: addressing data privacy and finding a balance between privacy and the need for more data; addressing IP issues; addressing AI bias; addressing challenges of automated AI content creation; providing guidance on how to develop trustworthy and explainable AI; providing more information on regulation frameworks and their impact; and offering more clear information and guideline about how to implement AI ethics in practice. In the following, we briefly summarise the various areas of AI research & application where more guidance is required, as highlighted by respondents in both groups:

- User data privacy concerns. How to make sure that data used in AI are authorised.
- AI regulation and ethics with regards to automatic content generation and use of synthetic content.
- How to ensure respect of the principles of inclusivity and non-discrimination in disseminated media content.
- How to harmonise the need for data to train AI models with the need for privacy. How to make sure that we know what kind of data are used in open datasets or retrained models. How to ensure that those that play by the rules are not put in disadvantage in the market by those going around regulation. Best practices on how to collect and use data.
- Guidelines or regulation with regard to content moderation.
- More information on media rights in relation to training sets (e.g. what images/videos are okay to use for training models). Possible licensing schemes for AI models respecting established copyrights (are existing data copyrights inherited to derived models?)
- AI regulation on user profiling and affective computing.
- Consideration of systems composed of multiple AI components rather than standalone system in regulatory efforts.
- Guidance on development of trustworthy and explainable AI solutions and relevant certifications.
- More information on the AI act and its impact on both AI research and media companies (for example, with regard to regulatory impact/exceptions regarding "media innovation involving AI" in media sandbox environments). There should be an effort to take care of ethics without blocking AI research or adoption of AI.
- How to implement/enforce in practice AI legal responsibility.

- Ways of guaranteeing freedom of speech and existence of alternative sources of information, as well as elevating the level of political, social, economic and cultural discourse by means of the new technologies.
- Creation of public online tools so that AI researchers can do self-assessments (e.g. for AI bias or for legal compliance).
- More focus on the environmental impacts of AI and how to make green AI.

It is important to point out that among media professionals, the majority of respondents request not just guidelines but clear-cut regulations for most of the main issues raised above. There is a similar trend also in the research community but it is not that accentuated. What is obvious from the above list of issues is that both the research community and the media industry are in need of clear guidance and much more information when it comes to the ethics of AI. People are struggling either to understand the impact of regulations or to address important issues such as data privacy and AI trustworthiness.

The second question is also an open question that aims to examine **what aspects of AI for media should be regulated** ("*What aspects of AI use in media do you think should be regulated or further regulated?*"). Again, the list is long but the most important issues highlighted by respondents include: user profiling and monitoring, data privacy, synthetic content generation, bias, targeted advertisement, AI transparency, and disinformation. The following list offers a more in depth look on the issues that respondents of the survey feel that need to be regulated or regulated more effectively:

- AI bias and discrimination against underrepresented or vulnerable groups. E.g. regulating the development and deployment of machine translation systems or decision-making systems.
- Respect of fundamental human rights (special mentions were made to children, with regard to social media use and impact).
- Impact of social media on young population (providing protection against depression, harassment, manipulation and isolation from the society).
- Use of / access to personal data.
- Data collection and processing of any kind.
- Deployment of user profiling/monitoring tools (including face recognition and emotion recognition systems and psychological profiling). One respondent made a special mention to cloud gaming where AI is able to provide very accurate and complete psychological profiles of the players.
- AI-controlled advertisement targeting users; recommendation systems.
- Creation and deployment of AI-synthesised or AI-manipulated content (video, audio, text). Clear labelling of such content should be required.
- Traceability of provided content (origin, right owners, distribution path).
- Manipulation of AI to influence public opinion – disinformation and deepfakes.
- Deployment of AI of questionable quality (e.g. trained on known biased datasets without applying methods to improve fairness, failing to reach stated objectives).
- AI explainability and transparency about how models are trained or decisions are made.

- Ensuring a level playing field between media content creators and platforms and dealing with big tech monopoly.

It is clear that the appetite for regulation that will, first of all, protect humans and, also, introduce clear rules about the development and deployment of AI (governed by the principles of trustworthy AI) is huge. People feel that issues like user profiling/monitoring, user targeting, illegal collection of personal data, generation of manipulated media content, and AI-enabled disinformation are getting out of control while they also increasingly worry about potential AI bias and discrimination or in general about AI that cannot be trusted. The EU should immediately address these fears, extending the dialogue with AI researchers, media industry, ethics and social science experts, and of course concerned citizens.

The third question examines what kind of measures different work environments have in place to **control the ethical risks of AI**. The respondents are asked to select more than one option from a list that contains measures like ethical AI principles, ethical AI checklist, following or being a member of a Code on AI ethics, ethical board committee, and ethics by design approaches. The results are visualised in Figure 31. As can be seen, 19% of the responders from the AI research community and 17% of media professionals do not even know whether their organisation has any processes or measures related to ethics management. In addition, 17% of AI researchers and 25% of media professionals state that they have no measures in place. These percentages show that lack of awareness or interest on ethics management (not only related to AI but in general) is a real problem both in the AI research community and also in the media industry. Among those who responded positively in the question whether they have any ethics related measures or processes in place the most common answer is ethical AI principles, followed by ethical board committees for AI researchers (mainly from academia and research centres) and ethical AI checklists for the media community. Only 13% of both communities is following or is a member of a Code on AI ethics. Finally ethics by design processes are followed by 23 % of the AI community and 19% of the media community. It is clear that much more work is required in order to raise awareness about the importance of ethics in both communities, especially in light of the AI breakthroughs that we expect to see in the next couple of decades.

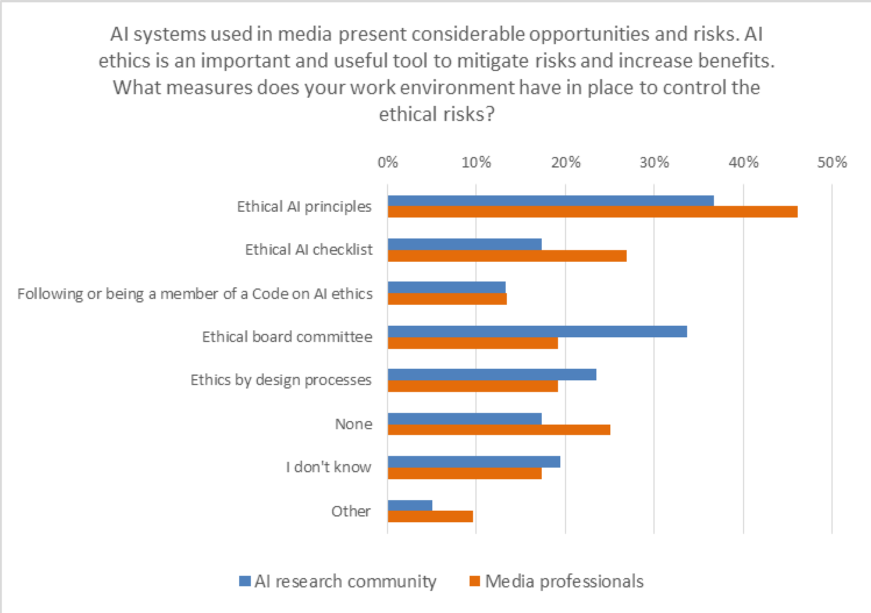


Figure 31: Workplace measures to control the ethical risks of AI.

The last question explores what kind of **impact assessments** are performed by media professionals and AI researchers. Three kinds of assessments are examined: data protection impact assessment (DPIA), human rights impact assessment (HRIA) and assessment list for trustworthy AI (ALTAI). The results are visualised in Figure 32. The most popular type of assessment are DPIAs with 36% of AI researchers and 19% of media professionals ever having done one. With regards to human rights assessment, the situation is even more disheartening: only 3% of AI researchers and 10% of media professionals has ever done one. The situation is similar for the assessment list for trustworthy AI (although kind of inversed): only 11% of AI researchers and 4% of media professionals has ever done one. DPIAs are now required based on the GDPR. This is perhaps the reason that DPIAs are much more common. As mentioned above, it is clear that more effort is necessary in order to raise awareness about the importance of ethics and these kinds of impact assessments in both communities.

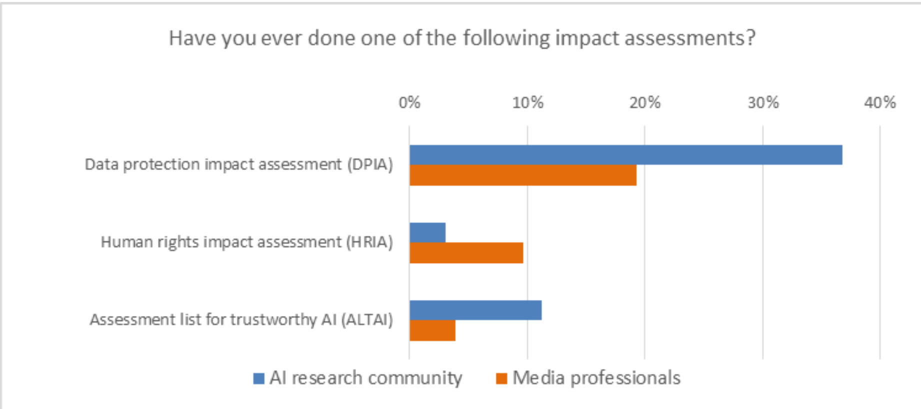


Figure 32: Impact assessments performed by AI researchers and media professionals.

4.2 AI4Media survey on Media AI in the service of Society & Democracy

In addition to the previous big-scale survey, which focused on collecting the opinions of media industry and AI research community representatives with regard to the use of AI in the media, a second small-scale short survey was also launched, aiming to shed light to the most important benefits and risks that the use of AI for the media brings to the society and democracy, also exploring ethical dimensions.

This survey was internal, addressed only to the AI4Media consortium, which comprises of members of both communities (AI and media). In total, 31 people completed the survey.

The survey included the following sections: (1) Professional background of survey respondents; (2) AI benefits for society & democracy; (3) AI risks for society & democracy; (4) Policies for ethical use of media AI & safeguarding of human rights. Sections 2 and 3 each included one multiple choice question + one complementary open-ended question (to elaborate, in case the previous choices were not enough) while section 4 included three multiple choice questions + one open-ended question. Answers to the question about the benefits of media AI for society and democracy were inspired by Khari Johnson's article "*How AI can empower communities and strengthen democracy*", published in VentureBeat[92]. Some screenshots from the survey questionnaire can be seen in Figure 33. The full version is available in the Appendix, in section 15.3.

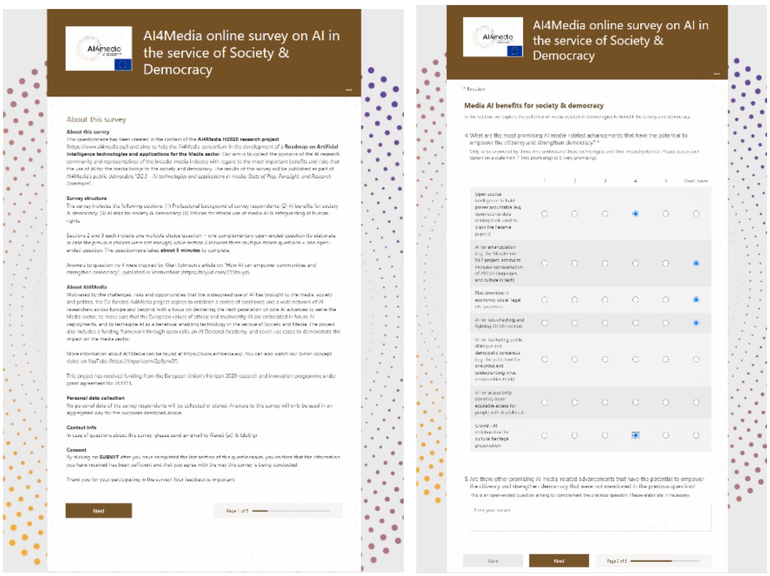


Figure 33: AI4Media internal survey on media AI in the service of society and democracy (full version available in the Appendix, section 15.3).

In the following subsections, we analyse the received responses. Each subsection corresponds to a different survey section.

[92] K. Johnson, "How AI can empower communities and strengthen democracy" (2020): https://venturebeat.com/2020/07/04/how-ai-can-empower-communities-and-strengthen-democracy/

4.2.1 Professional background of survey respondents

In total, 31 people completed the survey. 77% (24 respondents) belong to the research community while 23% (7 respondents) are from the extended media industry. With regard to the type of organisation they work for, 23 respondents work in research institutes or in academia (74%), one in the ICT industry, one in the media industry, 3 in PSM, and 3 in the public sector (see Figure 34).

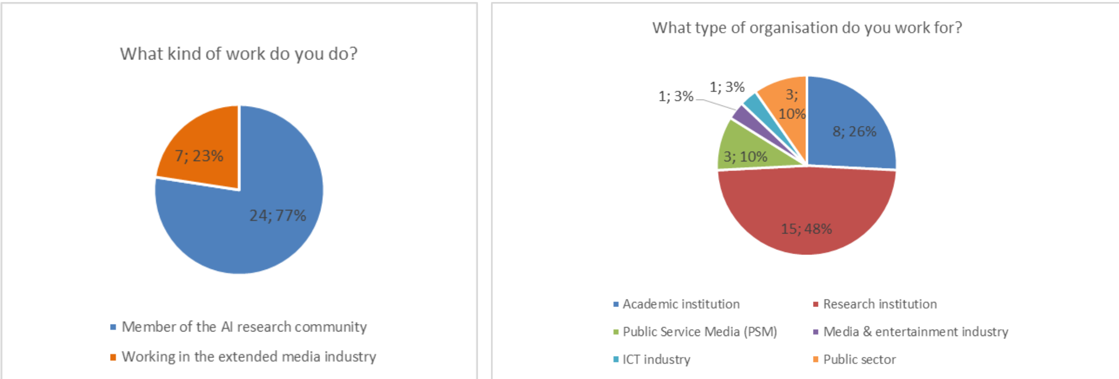


Figure 34: Information about the professional background of survey respondents.

4.2.2 Benefits of media AI for society & democracy

This section of the survey aims to explore the ***potential of media-related AI technologies to benefit the society and democracy***. Respondents were asked to assess a list of promising AI media-related advancements that have the potential to empower the citizenry and strengthen democracy, evaluating each item on the list on a Likert scale (1: not promising, 5: very promising). The list was inspired by Khari Johnson's article "*How AI can empower communities and strengthen democracy*" as mentioned above. We examine the following areas that could benefit from AI-enabled applications:

- Open source intelligence to hold power accountable (e.g. open-source data mining tools used to crack the Panama papers[93]);
- AI for emancipation (e.g. the Masakhane NLP project[94], aiming to increase representation of African languages and culture in tech);
- Bias detection in economic, social, legal etc. processes;
- AI for fact-checking and fighting disinformation;
- AI for facilitating public dialogue and democratic consensus (e.g. the pol.is[95] tool for analyzing and understanding what communities think);
- AI for accessibility (creating more equitable access for people with disabilities);
- Crowd - AI collaboration for cultural heritage preservation.

[93] M. Walker Guevara, How Artificial Intelligence Can Help Us Crack More Panama Papers Stories (2019): https://www.icij.org/inside-icij/2019/03/how-artificial-intelligence-can-help-us-crack-more-panama-papers-stories/
[94] Masakhane - A grassroots NLP community for Africa, by Africans: https://www.masakhane.io/
[95] Polis - Input Crowd, Output Meaning: https://pol.is/home

The results are visualised in Figure 35. As can be seen, all AI-enabled advancements are considered very promising (mostly, respondents answered with 4s and 5s). The most promising application of AI according to respondents is counteracting disinformation and fact-checking, followed by equitable access for all, and holding power accountable. The least promising seems crowd-AI collaboration for preserving the world's cultural heritage, perhaps because it is not still very clear to respondents how this can be practically done. The option for which we got the most "I don't know" answers was use of AI for the emancipation of underrepresented or minority groups. Again this is an area where much more research is necessary and projects like the Masakhane NLP project can show the way.

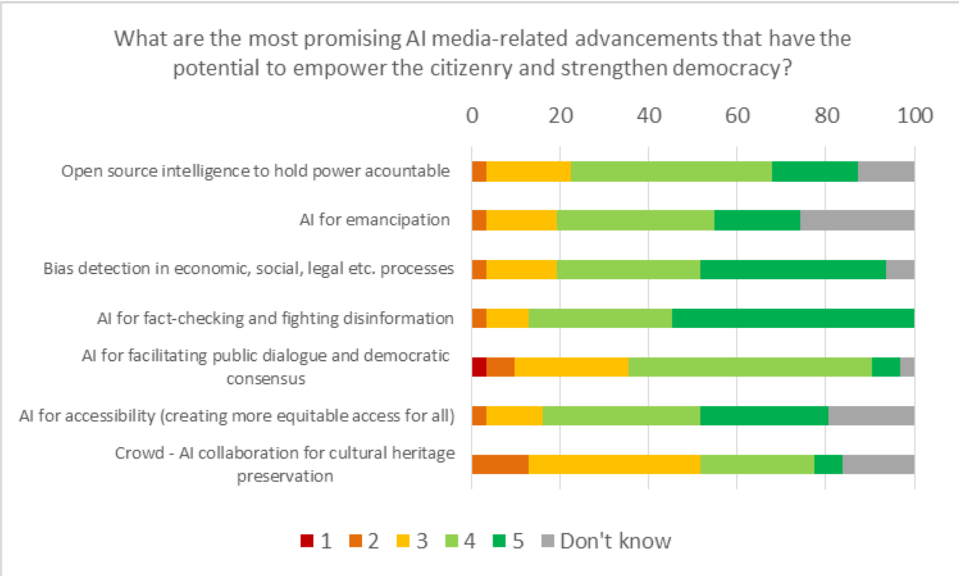


Figure 35: Promising AI media-related advancements that have the potential to empower the citizenry and strengthen democracy (Likert scale: 1 = not promising, 5 = very promising).

This section also included an open question, focusing on what other promising AI media-related advancements have the potential to empower the citizenry and strengthen democracy besides those included in the aforementioned list. The responses received focused on two distinct subjects:

- Diverse AI support for news media workflows, allowing journalists to focus on important tasks like investigative journalism, and production of semi or fully AI-generated media reports on niche topics/events, which would otherwise be impossible due to lack of resources.
- AI to support accessibility, not only for enabling access for people with disabilities but also for facilitating engagement and communication between people with different abilities and skills.

The first subject focuses on the potential of AI to create more informed citizens while the second focuses on smoothing communication barriers between different people.

4.2.3 Risks of media AI for society & democracy

This section of the survey aims to explore the ***risks of media-related AI for the society and democracy***. Respondents were asked to assess a list of media AI-related risks that have the potential to be detrimental to society and democracy, assessing each item on the list on a Likert scale (1: low risk, 5: high risk). The following risks have been included in the list, based on the analysis of sections 3 and 11 of this report:

- AI bias against groups of citizens, e.g. women, minorities, low-income people, vulnerable groups, etc.
- AI tools used to spread disinformation and fuel polarisation (e.g. bots, fake content generation);
- Over-personalisation of media services leading to filter bubbles;
- Use of AI tools to limit fundamental rights like freedom of expression;
- Unauthorised profiling of citizens;
- Unexplainable AI making decisions without human in the loop;
- AI affecting employment patterns;
- AI-induced inequality (people with access to data vs. people with no access).

The results are visualised in Figure 36. There are two risks that worry respondents the most: unauthorised profiling and monitoring of citizens (with 90% of respondents assigning to this risk a 4 or 5 in the aforementioned Likert scale) and use of AI to spread online disinformation and fuel polarisation on online discussions (with 93% of respondents assigning to this risk a 4 or 5). Disinformation is considered the number one risk of AI, followed by user profiling. Note also that there is not a single respondent that believes that these two are low-impact risks. The next most important risk is AI bias (with 77% of respondents assigning to this risk a 4 or 5), followed by over-personalisation of media services (with 71% of respondents assigning to this risk a 4 or 5). The risk that participants evaluate lower in this scale (receiving the most 1's and 2's, approx. 16% of all respondents) is unexplainable AI making decisions without a human in the loop. This finding may attributed to the fact that humans are very much still in the loop (and will probably be for a long time) and AI is not widely used for critical decisions that could have a detrimental effect on humans. The risk where the difference of opinions is higher is AI affecting employment patterns.

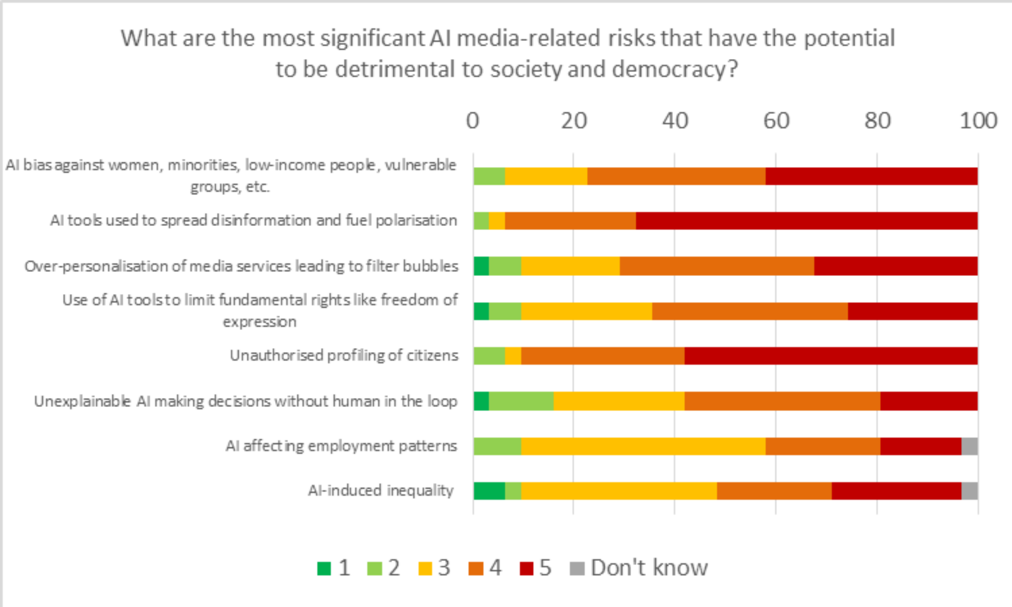


Figure 36: Media AI related risks that have the potential to be detrimental to society and democracy (Likert scale: 1 = low risk, 5 = high risk).

This section also included an open question, focusing on what other significant AI media-related risks that have the potential to be detrimental to society and democracy besides those included in the aforementioned list. The responses received focused on user monitoring, content and service labelling, and monopolies:

- User monitoring through hidden AI systems. Many citizens are unaware that AI tools are used in the services and products they use daily, hence they do not grasp their implications.
- AI-generated content that is not recognisable/labelled as such and is used for malicious purposes.
- Concentration of power via monopolies on data, computing resources and talent among a small number of big tech companies.

The need for labelling content generated by AI or labelling systems and services that use AI was also highlighted in the previous survey and would be a welcome step towards more AI transparency. Big tech monopoly of data and compute is another issue that has also concerned AI researchers and media industry companies, as showed in the first survey.

4.2.4 Policies for ethical use of media AI & safeguarding of human rights

The last section of the survey aims to explore policies to ensure ethical use of media AI and safeguard fundamental human rights. It includes four questions, which have been formulated by partner KUL, based on the analysis conducted in section 11 of this report.

The first question focuses on the **safeguarding of fundamental human rights** ("*In your opinion, what specific fundamental rights are not sufficiently safeguarded with regard to the use of AI in* *media applications?”).* The respondents were provided a list of rights and were asked to select the ones they thought were not efficiently safeguarded (more than one option could be selected). The results are visualised in Figure 37. As can be seen, respondents are mostly worried about the right to privacy and private life (74% of respondents), which is undoubtedly in danger due to online user profiling, monitoring and micro-targeting. The second concern involves the right to not be discriminated (65% of respondents), which may be in danger because of AI bias. Following relatively further behind is the right to freedom of expression and information (39% of respondents) and the right to property, including IPR (35% of respondents). With regard to the rest of the fundamental rights mentioned in the list, 23-26% of respondents worry about their safeguarding, with the exception of the right to freedom of assembly and of association, for which only 16% is worried. Finally, 10% of respondents answered “I don’t know”.

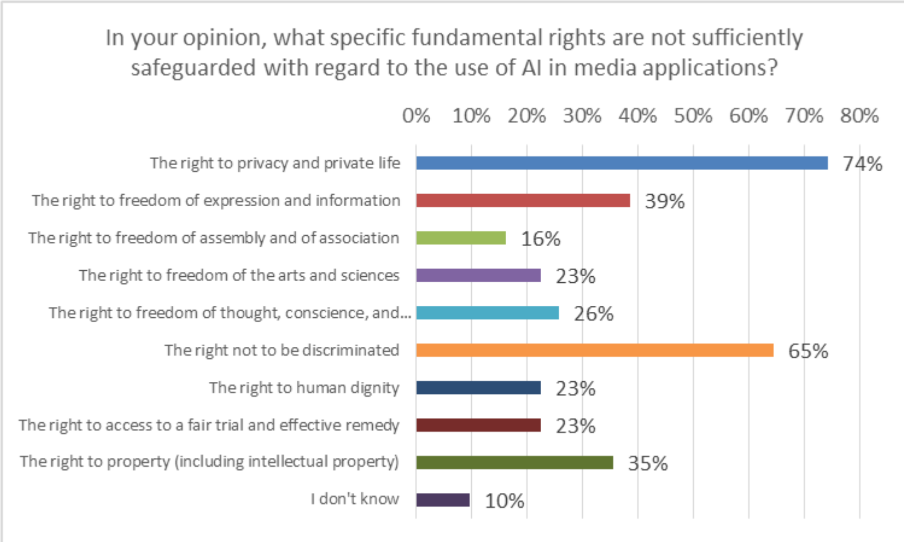


Figure 37: Safeguarding of fundamental human rights with regard to the use of AI in media applications.

The second question focuses on the **policies that should be adopted by governments and international organisations** to ensure that media AI is respectful of fundamental human rights (“*What kind of policies should States or international organisations put in place to ensure that AI media applications are respectful of fundamental rights?”).* The respondents were provided a list of policies and were asked to select the ones they thought should be adopted (more than one option could be selected). The following policies are included in the list:

- Regulations coming with fines for stakeholders developing or providing AI media applications (researchers, media companies, platforms);
- Independent authorities overseeing AI media applications' risks and opportunities and providing guidance;
- Standards, code of practices and other self-regulation instruments;
- Banning AI media applications when demonstrated to be considerably harmful to fundamental rights;

- Training and education programs on AI media applications and ethics / fundamental rights respect;
- Collaboration spaces for citizens, NGOs, media sector, ICT companies to discuss AI risks & mitigation measures.

The results are visualised in Figure 38. As can be seen, the main kind of policy respondents want to see from government and international organisations is oversight by independent authorities (77% of respondents). The policies that were least preferred are regulation with fines for stakeholders and ethics training (although they were still selected by 52% of respondents), probably for different reasons. The former finding is not surprising because regulation with fines is considered by many an aggressive policy on behalf of the government, which may hinder AI research or, on the other hand, create issues such as censoring. It is interesting however to notice that while only 52% of respondents believe regulation is a good policy, 65% believe that AI media applications should be banned if demonstrated to be considerably harmful to fundamental rights – banning being an extreme form of regulation. We have to note that all respondents selected at least two options, with almost half of them selecting all of them. Finally, while 29% of respondents believed some other type of policy is also necessary, only one provided an example, highlighting the need for more multidisciplinary approaches that involve AI and computer science as well as law, philosophy, humanities, etc. in education, training, and professional environments.

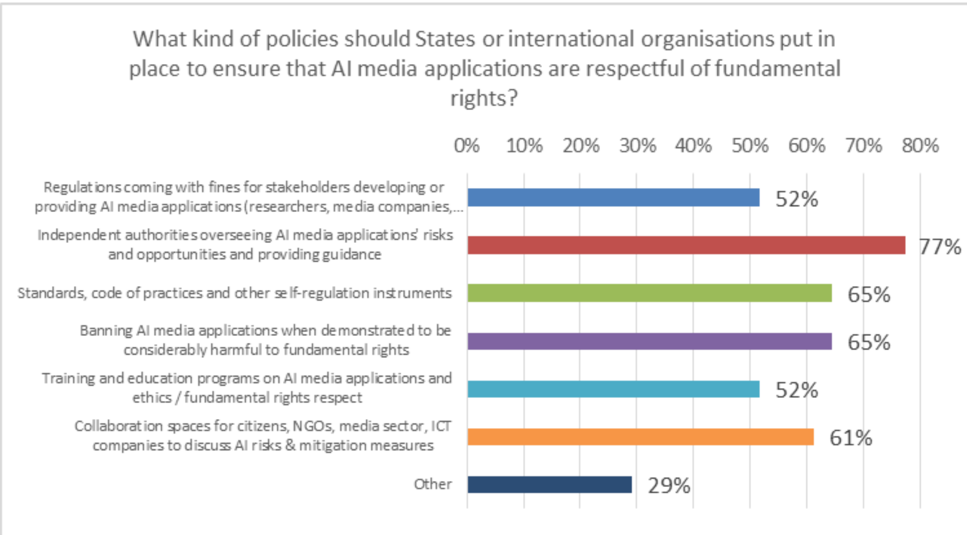


Figure 38: Policies that States or international organisations should put in place to ensure that AI media applications are respectful of fundamental rights.

The third question explored what types of **policies media companies should adopt** to ensure that media AI is respectful of fundamental human rights (“*What kind of policies should media companies put in place to ensure that AI media applications are respectful of fundamental rights?”).* The respondents were provided a list of policies and were asked to select the ones they thought should be adopted (more than one option could be selected). The following policies are included in the list:

- Reveal information on their algorithms and data;
- Publish yearly reports on how they use AI in media applications, including information about collected data/ethics handling/content recommendation/content moderation, etc.;
- Empower users to have influence on the content recommended to them, data collected from them, etc.;
- Ensure appeal rights and establish services for user questions and complaints, regarding the use of AI systems in media;
- Commit to and implement codes of conduct, certification, labels for compliant and ethical use of AI media applications;
- Enable independent research on their services to analyse potential impact and risks.

The results are visualised in Figure 39. As can be seen, what respondents mainly want to see media companies do, is to reveal information on their algorithms and data (77%). This has been a long-time request of NGOs and the civil society but also AI researchers. Other very popular policies include publishing yearly reports on issues with ethics repercussions (like handling of personal data or content moderation) (74%), empowering users to have influence on how the algorithms work with respect to them (71%) and independent research on the services of media companies to analyse potential risks (68%). The least popular policy, with less than half of the responders selecting it, is the establishment of appeal rights about media companies' decisions and services for user questions and complaints. This finding is a bit puzzling, given that already 71% of respondents agree that media companies should allow users to have control over their data, recommended content, etc. However, it could also be due to whether such policies (i.e. services that allow user complaints and demands) can be applicable at scale (and what resources would be required for that). Opinions are also split about ethical codes of conduct and AI certification (52%). This may be due to the experience of the last years (e.g. with the HLEG recommendation on disinformation) that shows that self-regulation and soft ethical codes may not be enough to actually ensure the ethical development and deployment of AI. One of the answers to the open question shed some extra light to this, at least with regard to the applicability of certifications by medium or small size companies.

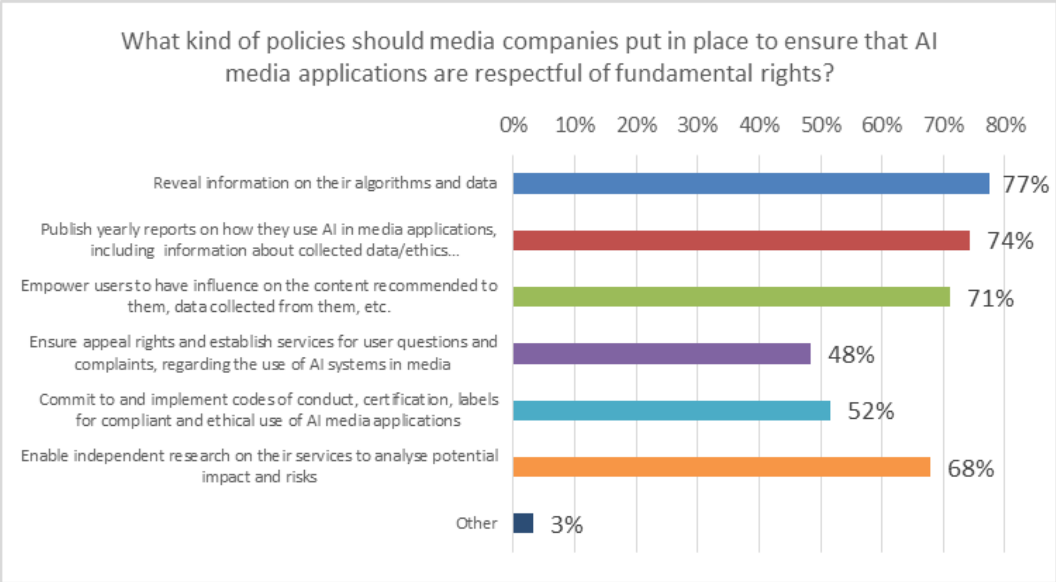


Figure 39: Policies that media companies should put in place to ensure that AI media applications are respectful of fundamental rights.

The final question of the survey is an open question that asks respondents to share advice for best practices and recommendations about AI media applications and fundamental rights ("*Do you have any best practices or recommendations to share regarding AI media applications respectful of fundamental rights?*").

One respondent points out that while all options about media company policies are useful, currently they are realistically applicable only on big media platforms like Facebook or Twitter that have the resources and capability to enforce them. However, in the case of media publishing companies, like BBC or similar organisations, some policies like AI certification are only realistic if there are in-house developed AI systems and if the instruments required to achieve the required AI transparency and trustworthiness are mature/usable.

Another interesting opinion provided highlights the need for processes that will allow implementation of these policies in practice. It is argued that there should be a bottom-up approach in formulating policy recommendations, engaging AI researchers, academics, media professionals, and citizens and including open public debates and forums. They also point out the need for efficient AI training from schools and academia to professionals in different industries and end users that it will help familiarise people with technological, ethical, and legal concepts around AI in order to be able to participate in the public debates.

4.3 Quick overview of survey results

Below, we briefly summarise the results of the two surveys, offering a quick overview of the main findings per survey section.

AI for the media industry

Future AI technology trends. Automatic content analysis & content creation, multi-lingual NLP, learning with limited data, explainable AI and fair AI are considered to be the AI technologies & applications with the biggest potential, aspiring to solve the most pressing problems of the media industry, which are: the ever-growing need for new content, the need to automatically analyse content to extract knowledge, the language barrier, the lack of data to train algorithms for new tasks or domains, and the need for AI that can be trusted.

Benefits of AI. Automation & optimisation of routine tasks, personalisation of content and services, increased productivity & operational efficiency, and enhancement of current services as considered to be the most important benefits that the use of AI brings to the media industry. Respondents were much more sceptical about the potential of AI to directly increase creativity. AI researchers were more eager to believe in AI's transformative role for the media than media professionals were, highlighting the need to increase awareness about AI's capabilities in media organisations and adopt clear strategies for its operationalisation.

Risks of AI. Unethical use of AI, biased AI, AI's lack of explainability, and AI used to hinder fundamental human rights are the risks that concern most survey respondents both as professionals but also as citizens. On the contrary, automation leading to loss of jobs or limiting human creativity is perceived as a lower risk, at least for now. At the same time, high expectations & low return, failure of AI in a critical mission, and ethical risks are the top reasons of concern for media industry companies.

Challenges of AI. AI researchers consider by far the greatest challenges the lack of data to train their algorithms but also the reluctance of the media industry to share their data to help and accelerate AI research. The US-China big tech monopoly in AI software and data, facilitated by enormous funding and generally much less regulation, is also something that creates challenges especially for European researchers. Media professionals, on the other hand, consider the integration of AI in business operations and processes, understanding what AI can do to improve workflows in the media industry, the lack of relevant skills in media professionals, and big-tech monopoly the biggest challenges for the wide adoption of AI by the industry.

AI strategies and skills. 73% of the PSM and the media & entertainment industry organisations affiliated with the survey respondents have invested in AI in the last year. Only 42% of the media professionals surveyed said that their organisation has an AI strategy in place. Among those organisations that do have a strategy in place, only 9% have a clear framework to manage AI-related risks (incl. ethical risks) and only 13% have established specific KPIs to measure impact of AI in the organisation. This lack of ethical AI strategies poses a significant risk for media organisations and public trust in media. AI software developers, data scientists, and experts on AI ethics/legal issues is the kind of skilled personnel that media organisations need in order to overcome the AI skills gap.

Adoption of AI solutions. Media professionals already use a variety of AI/ML applications, with the most popular ones being image/video analysis and NLP (with 44% each), followed by audio analysis (35%), and user experience/audience analysis (31%). Applications with low penetration are automatic-decision making for business (6%) and market analysis/forecasting (10%). Open source development tools, ready to use components from AI repositories, and co-development with research partners are considered to be the most promising ways for the development/adoption of AI tools in the media industry, showing the clear preference of respondents towards open AI solutions and collaborations that will lead to the adoption of AI that they can trust. Collaboration between the research and media communities (including shared projects, internships etc.) is considered the most promising approach to strengthen the transfer of AI knowledge and best practices from academia/research to the media industry. At the same time, researchers put special emphasis on sharing media industry data while media professionals point out the importance of training to increase the AI skills of media practitioners.

AI ethics & AI regulation. AI researchers and media professionals need more guidance by policy makers on implementation of AI ethics, highlighting topics such as data privacy (and balance between need for privacy and need for more data); IP issues; AI bias; automated AI content creation; development of trustworthy and explainable AI; and clear information on regulation frameworks and their impact. At the same time, they strongly believe stronger regulation is required for such issues (e.g. for user profiling and monitoring, data privacy, synthetic content generation, AI bias, targeted advertisement, AI transparency, and disinformation). The survey also showed the lack of ethical processes in many organisations, with almost 39% of all survey respondents declaring either that their organisation doesn't have specific ethical measures in place or that they do not know if it has. Similarly, only 11% of AI researchers and 4% of media professionals has ever done some kind of impact assessment like DPIA, HRIA, or ALTAI.

Media AI in the service of society and democracy

Benefits of media AI. The most promising application of AI that can benefit society and democracy according to respondents is counteracting disinformation, followed by enabling equitable access for all, and providing tools (e.g. for investigative journalism) to hold power accountable.

Risks of media AI. The risks of media-related AI that worry respondents the most with regard to potentially detrimental impact on society and democracy are unauthorised profiling & monitoring of citizens and use of AI to spread online disinformation & fuel polarisation on online discussions (with >90% of respondents saying that these are high or very high risks), followed by AI bias against groups of citizens.

Policies for ethical use of media AI. With regard to which fundamental human rights are not sufficiently safeguarded, respondents are mostly concerned about the right to privacy and private life (74%) and the right to not be discriminated (65%), which may be in danger due to online user profiling and monitoring and AI bias. In this direction, the main kind of policy respondents want to see from government and international organisations to ensure that media AI is respectful of fundamental human rights is oversight by independent authorities (77%), followed by codes of practices (65%), and banning AI media applications when demonstrated to be considerably harmful to fundamental rights (65%). The least preferred policy was regulation with fines for stakeholders (52%). With regard to the policies that media organisations should adopt, the most popular ones include: revealing information on their algorithms and data (77%), publishing yearly reports on issues with ethics repercussions (74%), empowering users to have influence on their data or how algorithms work (71%), and research by independent committees on the services of media companies to analyse potential risks (68%).

5 Core AI research: a glance into the future

In section 3, we analysed a selection of roadmaps, surveys, and articles focusing on media AI applications but also on more general AI technology trends. Based on the results of this analysis, we identified some of the most promising current and future trends in AI research, highlighting specific technologies that could benefit the extended media industry in section 3.3.

In this section, we further analyse some of these technologies, including *reinforcement learning, evolutionary learning, learning with scarce data*, *transformers*, *causal AI*, *AI at the edge*, *bioinspired learning* and *quantum computing*. The aim is to offer a clear overview of the current status of the technology, drivers and challenges for its development and adoption, future outlook in order to provide some information and insights on how these technologies could positively disrupt the media and entertainment industry.

A separate subsection is dedicated to each technology, including the following parts:

- ***Current status***, providing general information about the technology and its current status;
- ***Research challenges*** for the development and adoption of the technology;
- ***Societal and media industry drivers***, focusing on what potentially drives the adoption of this technology in the media and entertainment industry by providing ***vignettes***, i.e. short stories with media practitioners or users of media services, applications and content as the main characters that vividly describe how media professionals and users can benefit by the use of AI effectively, and how AI innovations could impact society and businesses;
- ***Future trends for the media sector***, presenting the potential applications of these technologies in different sectors of the media industry, explaining what kind of problems AI could help solve; and
- ***Goals for next 10 or 20 years***, summarising a set of milestones for the further development of these technologies and their application in the media.

This layout of the technology-focused subsections, and especially the *vignettes* and *future goals* parts, has been inspired by the "*20-Year Community Roadmap for Artificial Intelligence Research in the US*" report (see section 3.2.1).

5.1 Reinforcement learning

***Contributors**: David Melhart (modl.ai), Ahmed Khalifa (modl.ai), Daniele Gravina (modl.ai)*

Current status

Reinforcement Learning (RL) is an area of machine learning that is concerned with learning how to solve a problem through trial and error. For example, an agent can learn to drive a car by allowing it to control the car and providing it with positive or negative rewards based on how well it is performing (Figure 40). Until the recent advancement of machine learning, reinforcement learning applications were limited to toy problems because it was very hard to learn to perform well from just interacting with the environment. Nowadays, reinforcement learning is being used in a lot of different fields and showcases super-human performance across complex problems that humans are traditionally good at solving: those include playing popular board games like Go[96] and video games such as arcade-type Atari[97] and StarCraft II[98].

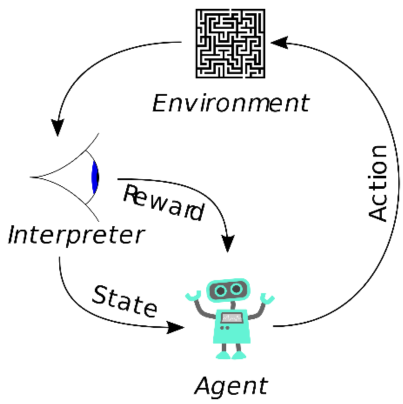


Figure 40: The components of the reinforcement learning framework.[99]

With the amazing performance of reinforcement learning algorithms over different games in the academic world, one might wonder if reinforcement learning is being used in other multimedia domains such as social media, online content production, movies or news outlets. Unfortunately, at the time of writing, RL is still not widespread in multimedia domains outside of video games. This is because although different types of multimedia provide a huge corpora of data where supervised and/or unsupervised learning can be used efficiently, traditionally RL is focused on controlling an agent in an interactive environment. However, looking at the recent advancements in reinforcement learning, one could see the advantage of using it in these fields, especially ***active learning***. Active learning algorithms involve agents that are continuously learning and adapting to an ever-changing situation. As such, active learning algorithms are well suited to address tasks involving human-computer interactions and domains where the focus is the preference or interaction style of humans using a system. For example, an RL bot could be used to assist directors during filming TV shows by learning their style of camera shooting over time and later propose camera positions that will fit the director's style and target aesthetics.

Research challenges

As mentioned before, RL is used more in games compared to other types of media. The lack of interactivity and the abundance of data usually prioritises different methods of machine learning over RL in other domains. Although RL is not a new field, its success is recent even within games.

---

[96] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." Nature (2016).
[97] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." NIPs (2013).
[98] Vinyals, Oriol, et al. "AlphaStar: Mastering the Real-Time Strategy Game StarCraft II." DeepMind Blog (2019).
[99] Image source: Wikipedia - https://en.wikipedia.org/wiki/Reinforcement_learning#/media/File:Reinforcement_learning_diagram.svg

Consequently, there are only a handful of examples of successful industrial applications of RL. An indicative example is the *Black & White* (EA, 2002) strategy game[100] and "god simulator". The game uses RL to control a creature's behaviour during the game that acts as the player's companion throughout the levels. The player can affect their creature's behaviour by observing their actions and giving positive or negative reinforcement quickly after. During the game, the creature can be trained to perform different tasks, aiding the player. Even though *Black & White* enjoyed universal acclaim at the time, not a lot of companies followed suit in implementing RL algorithms in their games. The prototypical nature of the algorithm in game development did not help RL to advance further and become mainstream in games and their design.

Another challenge faced by RL is that trained agents may reach a *super-human performance*[101,102,103]. While desirable in other domains, superhuman players are not an interesting feature in games because humans either won't be able to compete with these AI agents (if AI is an adversary) or they will be outshined by the agent which might make the game boring (if AI has a companion role). Besides that, designing these AI systems for any media sector is a complex task as there are no standard libraries that can be easily integrated and used out of the box. Finally, training these AI agents is often costly both in terms of money and time[104]. That cost, in turn, does not allow for fast iterations and quick testing of new ways of using RL in different media sectors.

Although reinforcement learning reached super-human performance in many fields, it still faces several challenges with *generalisation*. Generalisation is the problem of using the trained agent on a new problem other than the one it got trained on. For example, if we trained an agent to play *Super Mario Bros* (Nintendo, 1985)[105] optimally, we would like it to be able to do so in *Super Mario Bros 3* (Nintendo, 1988) too. Overcoming the generalisation challenge will help game and media productions reduce their overall cost as AI engineers won't need to retrain AI models across different games, similar to how computer vision models can operate directly without pre-training across dissimilar pattern recognition tasks[106]. We argue that transferable and generalisable world models and agents will be available in games and other media fields as soon as such AI models become easier to integrate, more accessible, and cheaper to train.

Societal and media industry drivers

Vignette: AI bots for game development & quality assurance

Gita and Rami are a programmer and a quality assurance (QA) manager working at a medium-size game company. Their company is working on a multiplayer action-adventure game "*Power of Legends*", which will be supported by continuous small content updates throughout the whole life-cycle of the game. Dependence on regular content updates means development never ends.

[100] Black and White Video games, EA (2022): https://www.ea.com/games/black-and-white
[101] Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." Nature (2016).
[102] Mnih, Volodymyr, et al. "Playing atari with deep reinforcement learning." NIPs (2013).
[103] Vinyals, Oriol, et al. "AlphaStar: Mastering the Real-Time Strategy Game StarCraft II." DeepMind Blog (2019).
[104] Wang, Ken. "DeepMind achieved StarCraft II GrandMaster Level, but at what cost?." Medium (2020)
[105] Super Mario Bros, Nintendo (1985): https://mario.nintendo.com/history/
[106] Shah, Hardik. "How Artificial Intelligence is Enhancing Mobile App Technology." RTInsights (2020).

Content must be added at speed while maintaining quality and creativity to keep players engaged and attract new ones, which means that the average cost of game development has increased ten folds each decade so far. To mitigate these costs, Gita and Rami's company uses AI-assisted testing tools that leverage reinforcement learning to quickly train agents that can perform the more menial tasks of game testing. This tool helps both the programmers and quality assurance staff to manage the rapidly increasing testing demand.

In the new content upgrade to "*Power of Legends*,'' the developers are introducing new maps and items to the game. The team has to make sure that the new map is bug-free and playable and that the newly added items do not break already existing levels. The work starts with Gita, who is responsible for the integration between the game and the testing software. She sets up the AI-powered bot in the game environment by defining its input space and the basic glitch or bug criteria. The goal of the setup is to prepare the work of QA staff, who will have a higher level control over the bot's interface. Once the bot is integrated into the game, the QA staff does not need to pull resources from the programmers; they can modify the bot's behaviour on their own to run a plethora of different tests. While not completely replacing traditional quality assurance, using AI-based game testing can deliver a faster, cheaper, and more consistent turn-around.

Once a build of the game is prepared for testing, the game is uploaded into the testing platform, where a QA session can be initiated. At this point, Rami takes over and starts a test through a graphical user interface. The AI bots adapt to new content with ease (Figure 41) and quickly find bugs and glitches with different levels of severity. Some of these are game-breaking issues that have to be prioritised and addressed immediately by the programmers, while others are minor glitches or events that break the game's balance. Once the bot finishes, Rami receives a detailed test report which includes a smart summary of the encountered issues and a structured telemetry log of these issues. The created bug reports are sorted based on severity and sent back to Gita, who can immediately start working on an update, giving Rami time to go through the telemetry logs and - if needed - replay or reproduce more complex issues. This faster iteration cycle between Gita and Rami aided by the AI bots frees up resources in both the development and testing process. As the AI bots work faster and more consistently than human testers, there is less idle time during development. Gita can deliver faster updates to the game and Rami can focus on parts of the QA process that require more insight.

Recently, large-scale game developer studios started looking into using reinforcement learning in the game development process instead of using it to control characters in the game. The superhuman performance[104,106,107] and fast exploration power allowed these companies to utilise these agents to test their games. For example, EA uses reinforcement learning to detect navigation problems such as unreachable areas or missing collision boxes[108,109]. Companies like modl.ai are working on creating plug-and-play AI tools, which can integrate into different games,

[107] Koster, Ralph. “The cost of games” Venture Beat (2018).
[108] Thompson, Tommy. “Training the Shadow AI of Killer Instinct.” AI and Games (2017).
[109] Yannakakis, Georgios N., Antonios Liapis, and Constantine Alexopoulos. "Mixed-initiative co-creativity." Foundations of Digital Games (2014).

which could help bring these innovations to small and mid-sized companies as well, making Gita and Rami's story a reality.

Future trends for the media sector

The use of RL will increase in the game industry and other media sectors at large over the next decade. In the game industry, the use of reinforcement learning in the area of ***automatic game testing*** is going to grow more over the next decade and receive much research and innovation attention (Figure 41). This increase of interest in RL-based game testing is due to games gradually becoming online services (for instance *Fortnite*, *Dota 2*, *Candy Crush*, etc.) that require continuous updates and new content to keep players engaged. To keep that supply running, any new updates and content need to be tested periodically, resulting in a substantial increase in the production cost of a game. This cost increase is far from being sustainable. Thus, if game companies wish to scale their productions and expand their online services, they will need to rely on reliable and explainable AI methods that are deeply integrated into the testing process.

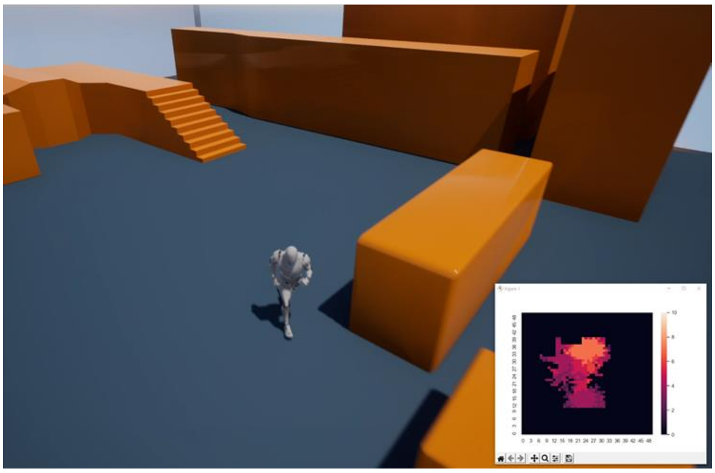


Figure 41: Modl.ai AI bot exploring a previously unseen game level.

Besides games, one might think about the various ways RL can be used in other domains throughout other media sectors. For instance, it could be used to create ***new types of media content*** such as (parts of) games, levels, environments, visuals, cloth simulations for fashion design, camera positions during filming movies and TV shows, etc. These avenues are not explored yet in any sector of the media industry but we expect that recent advancements in machine learning and RL will push innovation in these fields as well. But to see this happening, content creators should always keep their eyes and minds open to embrace these super-human performance agents and find new ways to incorporate RL in different media sectors. For example, the EA research team has been using RL recently to generate new playable levels for different games with different levels of difficulty[110] . Such generative AI methods could be used

[110] Gisslén, Linus, et al. "Adversarial reinforcement learning for procedural content generation." IEEE Conference on Games (CoG) (2021).

in the other media sectors as well for creating/curating content that will most likely be enjoyed by their users.

In addition to RL, ***imitation learning (IL)*** has been gaining popularity in recent years. Imitation learning is an area of machine learning concerned with the problem of learning how to imitate humans through their behavioural data (also called human demonstrations). The reason for their popularity is that IL algorithms may yield agents that perform and behave in a human-like fashion. This feature is important in competitive multiplayer games such as *Fortnite* (Epic, 2017)[111] and *Dota 2* (Valve, 2013)[112] as in these games players usually need believable and human-standard AI to play with or against. Most scripted (not machine-learned) AI agents do not perform as well as humans and they tend to follow specific strategies which can be exploited easily by players. Through imitation learning, we make multiplayer games engaging and plausible even when human players are not always there. By using the collected player data we can train AI agents that capture a player's style and performance which we can, in turn, use to create NPCs in the game for humans to play against (or with). *Forza* (Microsoft, 2018)[113] and *Killer Instinct* (Rare, 2013)[114], for instance, use the player's data to create a shadow player (driver and fighter, respectively) that other players can play against and test their skills[108,115]. Most notably the *Drivatar AI* system of the Forza Series (2005-present) is the longest living imitation learning algorithm existent in video games with more than 15 years of continuous development and operation.

Goals for next 10 or 20 years

We expect more media and game companies to start adopting RL in all stages of media production from early design to shipping. In the next 10 years, RL algorithms will not be used as a fully autonomous system but it will be working in tandem with designers, developers, content creators, and testers in a mixed-initiative fashion[116]. These tools will provide designers with new ways of exploring new rules and mechanics, help them design levels, test the designed content, and provide statistics on the user experience. They will elevate the tester's job to focus more on high-level challenges instead of tedious repetitive work. They will also help developers to optimise their code and find better solutions toward simulating certain physics or world history. They will help content creators in the content creation process by suggesting improvements and edits and sometimes whole topics that consumers are looking for. Because of all the aforementioned reasons, content creation cycles will start to be shorter which will not only cut the cost of production but also it will allow the industry to be more experimental and take more risks.

---

[111] Fortnite, Epic: https://www.epicgames.com/fortnite/en-US/home
[112] Dota 2, Valve: https://www.dota2.com/home
[113] Forza, Microsoft: https://www.microsoft.com/en-us/store/collections/forzacollection
[114] Killer Instinct, Rare: http://www.raregamer.co.uk/games/killer-instinct/
[115] T. Thompson, "How Forza's Drivatar Actually Works." AI and Games (2021).
[116] Yannakakis, Georgios N., Antonios Liapis, and Constantine Alexopoulos. "Mixed-initiative co-creativity." Foundations of Digital Games (2014).

To achieve the long-term goal of wide AI (RL) adoption in media and game productions certain decisive steps need to be made towards rapid deployment and wide accessibility. Most notably, the Unity game engine has been addressing the accessibility issue since 2018: it released an easy to integrate library called ML-Agents[117]. This library allows Unity developers to train agents using reinforcement learning in their own games. Although this is an important first step towards making reinforcement learning/machine learning more accessible, the algorithms are still computationally heavy and hard to debug. Besides, ML-Agents are currently operational only within Unity which is far from being ideal for game developers that use other engines.

We think that in the next 20 years, easy-to-use RL libraries will be released (similar to ML-Agents). These libraries will allow non-technical users to easily interact with them and train them on more specific problems by a press of a button. Besides that, the advancements in hardware and software will enable RL algorithms to run efficiently on small devices like smartphones, which will allow users to have the power of these models at the tip of their fingers.

5.2 Evolutionary learning

***Contributors**: Georgios N. Yannakakis (UM), Antonios Liapis (UM)*

Current status

As its name suggests, ***artificial evolution*** emulates the paradigm of evolution of the species, put forward in the 18th century by biologists such as Darwin and Lamarck, as a way to find good solutions when given specific goals and constraints. At the highest conceptual level, artificial evolution creates a population of initial solutions, evaluates how good they are in terms of an objective, then selects the best among them and stochastically adjusts their parameters (often recombining two or more solutions together) to create a new population of offspring. This evaluation, selection, genetic change and reinsertion (see Figure 42) is carried out over multiple generations until the population converges towards better solutions.

Evolutionary computation has over 50 years of history[118] and is one of the pillars of computational intelligence (along with machine learning and reinforcement learning). In terms of the problems that artificial evolution is often called to solve, extensive focus has been placed on evolving computer programs that are able to carry out computations with a fairly freeform underlying structure, as well as numerical optimisation tasks where the potential of artificial evolution to reach global optimal solutions is most advantageous. Unlike machine learning, artificial evolution explores the search space stochastically, often maintaining a population of solutions.

[117] A. Juliani, et al. "Unity: A general platform for intelligent agents." arXiv preprint arXiv:1809.02627 (2018).
[118] K. A. De Jong. Evolutionary Computation: A Unified Approach. MIT Press, Cambridge, MA, USA. 2006

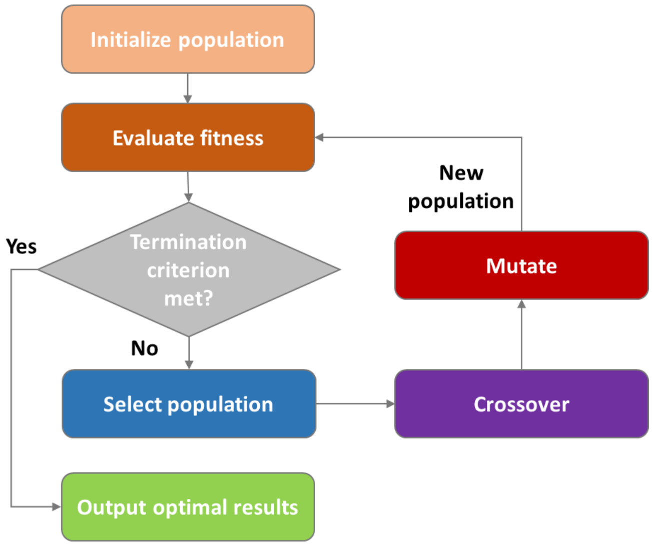


Figure 42: Evolutionary learning general concept[119].

In the media sector, artificial evolution has often been used to automatically or semi-automatically generate ***computational art***. Some of the earliest instances of evolutionary art were the line drawings of Dawkin's *Biomorph*[120] (Figure 43) and Karl Sims' evolved computer programs that could generate 2D or 3D plants or 2D images[121]. The stochastic nature of artificial evolution, which allows for more wide exploration of the solution space compared to gradient descent, has made them particularly popular for the research of computational creativity[122].

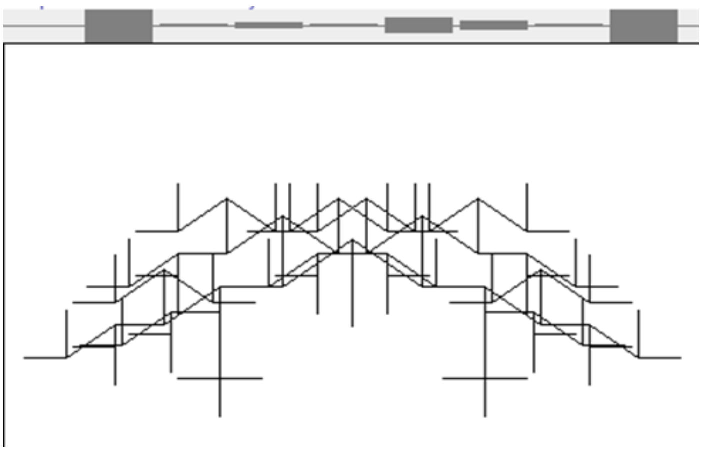


Figure 43: Dawkin's Biomorph[123], arguably the first instance of evolutionary art.

[119] Figure adapted from A. Doku, FitJSP - Fancy Interactive tool for Job-Scheduling problems (2020): https://blog.arinti.be/fitjsp-fancy-interactive-tool-for-job-scheduling-problems-791a9f6453ff
[120] R. Dawkins. The blind watchmaker: Why the evidence of evolution reveals a universe without design. W.W. Norton & Co, New York, NY, USA. 1987.
[121] K. Sims. Artificial evolution for computer graphics. In Proceedings of the 18th SIGGRAPH. Association for Computing Machinery, New York, NY, USA, 1991.
[122] S. Colton, R. Lopez de Mantaras & O. Stock. Computational Creativity: Coming of Age. AI Magazine, 30(3), 11. 2009.
[123] Image source: Wikipedia - https://en.wikipedia.org/wiki/File:BiomorphBounce.png

Computational creativity is “*the philosophy, science and engineering of computational systems which, by taking on particular responsibilities, exhibit behaviours that unbiased observers would deem to be creative*”[122]. Research in computational creativity often follows the paradigms established around human creativity, such as the concept of “p-creativity” (***psychological creativity***), when a creator considers their creations novel and valuable regardless of whether others would agree, and “h-creativity” (***historical creativity***), which introduces previously unimagined ideas or inventions into the world. Applications of computational creativity focus on artistic expression, including visual art (see for example Figure 43, Figure 44 and Figure 45), music, narrative, humor and poetry[124]. Unlike numerical optimisation, however, formulating what constitutes “quality”, “novelty” or a “valid solution” in computational creativity and evolutionary art raises significant challenges and debates[125].

Figure 44: PicBreeder[126] uses neuroevolution to produce images that multiple users can interact with, evolve further, and evaluate through a public website.

When applying evolutionary computation to the media sector and towards computational creativity more broadly, a major challenge is assessing the quality of generated content in such aesthetic-oriented, subjective domains, in order e.g. to guide the generator towards better content. It is not surprising that some of the early work in evolutionary art rely on a human

[124] S. Colton & G. Wiggins. Computational creativity: The final frontier?. Frontiers in Artificial Intelligence and Applications. 2012.
[125] G. Ritchie. Some empirical criteria for attributing creativity to a computer program. Minds and Machines, 17:76–99, 2007
[126] J. Secretan, N. Beato, D.B. D'Ambrosio, A. Rodriguez, A. Campbell, J. T. Folsom-Kovarik, and K. O. Stanley. 2011. Picbreeder: A case study in collaborative evolutionary exploration of design space. Evol. Comput. 19, 3 (Fall 2011), 373–403. Image source: http://picbreeder.org/ (accessed 15 Dec. 2021)

curator to guide evolution; similar practices in leveraging humans to perform interactive evolution are still popular today (an example can be seen in Figure 44).

Research challenges

A core challenge of applying evolutionary computation in the media sector remains the ***evaluation of quality***. To guide artificial evolution towards better content, it is common to use existing corpora such as human ratings[127], classifiers between man-made and generated images[128], object recognition[129] (see also Figure 45) or models that match images with language[130].

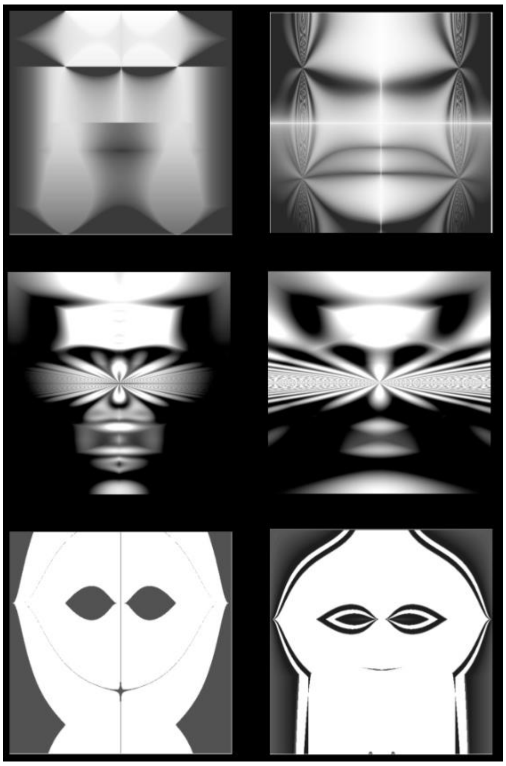


Figure 45: Evolved grayscale images guided by an evaluation based on the certainty of a trained classifier of human faces[131].

[127] S. Baluja, D. Pomerleau, and T. Jochem. Towards automated artificial evolution for computer-generated images. Musical networks, pages 341–370, 1999

[128] P. Machado, J. Romero, A. Santos, A. Cardoso, and A. Pazos. On the development of evolutionary artificial artists. Computers & Graphics, 31(6):818–826, 2007.

[129] J. Correia, P. Machado, J. Romero, and A. Carballal. Evolving figurative images using expression-based evolutionary art. In Proceedings of the International Conference on Computational Creativity, pages 24–31, 2013.

[130] D. Norton, H. Darrell, and D. Ventura. Establishing appreciation in a creative system. In Proceedings of the International Conference Computational Creativity, pages 26–35, 2010.

[131] Image source: P. Machado. 2021. Evolutionary art and design: representation, fitness and interaction. Proceedings of the Genetic and Evolutionary Computation Conference Companion. Association for Computing Machinery, New York, NY, USA, 1002–1031.

Treating evolutionary search as an *optimiser*, when it comes to creative media content generation, can be limiting and short-sighted. It is arguably impossible to adequately address the problem of searching for a "best" solution in domains and problem spaces where "good" is a subjective notion, as well as one that is deeply related to the context of use, intent, or current trends in a community of human (or AI) artists. Searching for solutions that are different from each other[132] can somehow mitigate this issue, by attempting to explore the space as thoroughly as possible rather than towards short-term exploitation. In experiments with robotics, divergent search has shown to be efficient in handling deceptive problems, where the final goal can only be reached by passing through "bad" parts of the space according to a pre-constructed, quantifiable notion of goodness[132]. However, what constitutes novelty in divergent search algorithms, as well as how quality can be formulated and maintained in quality-diversity algorithms remain open research challenges that can be as difficult to tackle as formulating an objective function for such subjective domains.

Societal and media industry drivers

Vignette: Designing new levels in a video game

Kiko is a game developer who wishes to design a new level for their upcoming game "*The Fabulous Journey*". Kiko works at a small game studio with only three developers (one programmer, one artist, and one level designer). Due to their small team, if they wish to publish "*The Fabulous Journey*" within the next two years they must rely on procedural content generation. Therefore, "*The Fabulous Journey*" is a series of spatial challenges and Kiko is tasked with designing these individual levels taking advantage of the power of the AI tool at their disposal. This puts low demands on the artist with regards to creation of art assets, character designs and animations since content is re-used throughout the levels, while the programmer can focus on the gameplay mechanics that are persistent throughout all levels. Kiko sits down to come up with an idea for the next level, starting up the AI tool and loading Kiko's profile, which contains all the history of their interaction and past designed and/or generated levels. Kiko considers some of the planned mechanics that the programmer has suggested, and wants to highlight a "wall-jump" mechanic that has not been very often used so far in past levels. The programmer has already updated the AI tool to use the current version of the game which uses the new mechanic, and has added a logger to count uses of "wall-jump" during a play session. Kiko uses the graphic user interface of the AI tool to specify the constraints for the next task: "use [wall-jump] [at least] [5] times" (text in brackets are options in drop-down lists). Kiko also uses the graphic user interface to specify that they want to explore levels of different length, and with different numbers of enemies. Kiko could also change what would qualify as the "best" level, but they keep the same metric that the studio has been using throughout production, which is the level that has the highest score after a simulation with an AI agent.

With everything setup, Kiko presses the *Generate* button and after a short while, a number of level layouts start appearing on the screen. Kiko can wait until the system has produced all the

[132] J. Lehman and K. O. Stanley, "Abandoning objectives: Evolution through the search for novelty alone," Evolutionary computation, vol. 19, no. 2, 2011

best levels with few enemies, many enemies, short length, long length and any combination thereof. Kiko can choose the computer-generated levels, and get summary statistics from the simulation (such as score of the agent at the end of the level, number of times that each mechanic was used, number of deaths by enemies, number of enemies killed etc.). Moreover, Kiko can choose to watch the AI agent's playthrough by re-running the simulation. Finally, Kiko can choose to play through the level themselves. Kiko can also stop the AI generation earlier, and select some of the levels that they prefer. Then Kiko can continue running the automated process: the AI will generate levels more similar to Kiko's selections. Alternatively, Kiko can enter an editing tool and directly modify the levels generated by the AI. Once Kiko has finished editing, they can playtest the level themselves or have an AI agent playtest it and report some summary statistics of the simulation. Kiko can export the levels they created or some hand-picked generated levels, and add them to the current version of the game. The programmer, artist, and also Kiko can further edit these in future iterations (e.g. after more content or more code has been produced) with or without the use of the AI tool.

Future trends for the media sector

Evolutionary computation is a powerful tool for exploring a large variety of designs. Coupled with other computational intelligence algorithms, such as recommender systems and latent representations, evolutionary computation can produce high-quality and personalised content that is appropriate to show to a designer during their workflow. Often, such AI-provided ideation mechanisms are used in early conceptual phases, allowing the designer to see many options but leaving more room for human creativity and control during later stages of the design where getting the details right is critical. Moreover, allowing the designer to keep their own artistic vision of the final product is imperative: this can be facilitated by interfaces that customise the initial parameters for exploration, by choosing specific examples that the user would prefer the AI to move towards, or by manually editing interim products to help the AI start from a better seed and move towards a better direction. All of these modes of interaction are described in the vignette above.

Based on these properties and requirements for integrating evolutionary algorithms in the design process, future research trends in this vein for the media sector will have to address four main issues: (a) the ***type of representations*** that evolution can explore, (b) the way in which ***quality and diversity*** are calculated, (c) ways of ***modelling designers*** in order for the AI to produce more personalised artefacts, and (d) ***interaction paradigms*** for the users to be able to view, control, and make use of the generated artefacts. The first two goals are likely best tackled through deep learning methodologies which can produce a more compact representation that evolution can more easily explore (compared to e.g. pixel-level representations), while supervised and unsupervised learning can be used to train predictive models of an artifact's quality and a dataset's diversity respectively. These predictive measures of quality and diversity can be used instead of the current mathematically defined formulas for quality-diversity evolutionary algorithms. Early work has already started to explore this direction, in non-media domains[133].

The last two goals are as reliant on traditional human-computer interaction and user modelling as they are on explainable AI research[134]. Research in this vein has so far focused on which of the large number of generated content to show to a user, given the cognitive overload of too many options. Moreover, early work has focused on personalising designer models in terms of different prescriptive quality and diversity dimensions, based on which AI generated solutions the user tends to select over unselected ones[135]. However, with the introduction of machine learning in all aspects of the design pipeline (including the future trends of combining evolution and latent vector representations discussed above), explaining to the designer why the level is considered of good quality – or more importantly, why the system considers the level appropriate for this designer – is vital for the designer to be able to trust the system and use it for their creative work. Especially in domains where artistic vision and creativity are paramount, the only opportunity for AI to be able to prompt co-creativity[136] is by providing ***an AI partner that can be both supportive but also "honest" and transparent.***

Goals for next 10 or 20 years

Public and academic attention in generative art has been largely driven by deep-learning-based architectures. One can only expect that short-term future accomplishments in creative domains such as images, music, and text generation will largely rely on these corpora-driven, trained models and AI methods such as transformers[137], generative adversarial networks[138] and style transfer[139]. However, ensuring that evaluation of the quality of generated media is scientifically robust and replicable remains an open challenge that will need to be addressed in the next 10 years. Theoretical constructs from computational creativity research[125] can shed important light on identifying the novelty and quality of such generated artworks. Standards for what constitutes original and "authentic" are necessary not only for the purposes of ascertaining intent and creativity but also for handling Intellectual Property and financial gain.

---

133 A. Cully. Autonomous skill discovery with quality-diversity and unsupervised descriptors. In Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '19). Association for Computing Machinery, New York, NY, USA, 81–89. 2019

134 J. Zhu, A. Liapis, S. Risi, R. Bidarra and G. M. Youngblood, "Explainable AI for Designers: A Human-Centered Perspective on Mixed-Initiative Co-Creation," Proceedings of the IEEE Conference on Computational Intelligence and Games, 2018.

135 A. Liapis, G. N. Yannakakis and J. Togelius. Designer modeling for Sentient Sketchbook. Proceedings of the 2014 IEEE Conference on Computational Intelligence and Games, 2014.

136 G. N. Yannakakis, A. Liapis and C. Alexopoulos. Mixed-Initiative Co-Creativity. Proceedings of the 9th Conference on the Foundations of Digital Games. 2014

137 T. Brown, et. al. Language Models are Few-Shot Learners. Advances in Neural Information Processing Systems 33. 2020.

138 I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial networks. In Proceedings of the International Conference on Neural Information Processing Systems, 2014

139 Z. Hu, J. Jia, B. Liu, Y. Bu, and J. Fu. Aesthetic-Aware Image Style Transfer. In Proceedings of the 28th ACM International Conference on Multimedia. Association for Computing Machinery, New York, NY, USA, 3320–3329. 2020.

Beyond short-term accomplishments, integrating deep learning paradigms in interactive evolution is expected to lead to more promising long-term accomplishments. The main benefit of artificial evolution and specifically divergent or quality-diversity search would be that it explores the possible set of solutions better, and can thus lead to more varied outcomes than the gradient-based methods used currently. A major drawback of evolutionary computation is that this additional exploration comes at a computational cost. Long-term hardware and engineering developments that can parallelise such search processes can benefit the media sector both by training better deep learning models through neuroevolution[140] and by generating a more diverse set of creative artefacts based on pre-trained representations[141] and driven by pre-trained or custom-trained evaluations of quality, diversity, appropriateness, personal preference, and more.

5.3 Learning with scarce data

Contributors: *Giuseppe Amato (CNR), Alejandro Moreo Fernandez (CNR), Hannes Fassold (JR), Werner Bailer (JR), Mihai Dogariu (UPB)*

Current status

The great success achieved during the last decade by deep learning based algorithms has been supported simultaneously by the availability of massive amounts of training data, and high performance GPU powered computing resources. However, there are applications where there is a lack of high quality annotated training sets and performance of AI algorithms in these scenarios is still a challenging issue.

Having a dataset of insufficient size for training usually leads to a model which is prone to overfitting and performs poorly in practice. In many real-world applications based on multimedia content analysis, it is simply not possible or not viable to gather and annotate such a large training dataset. This may be due to the prohibitive cost of human annotation, ownership/copyright issues of the data, or simply not having enough media content of a certain kind available.

This still open problem has been addressed so far using various solutions, which can be roughly classified into the following categories:

Transfer learning: Transfer learning aims at exploiting knowledge acquired while addressing a problem to solve a related but different problem[142] (Figure 46). In the deep learning scenario, this is generally obtained by using a pre-trained model (a deep neural network trained in a

[140] K. O. Stanley and R. Miikkulainen. Evolving neural networks through augmenting topologies. Evol. Comput. 10, 2 (Summer 2002), 99–127. 2002.

[141] M. C. Fontaine and S. Nikolaidis. Differentiable Quality Diversity. Proceedings of the 35th Conference on Neural Information Processing Systems. 2021.

[142] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, Qing He, A Comprehensive Survey on Transfer Learning, arXiv:1911.02685v3, 2020

certain applicative scenario) to address a new application. For instance, in a multimedia retrieval setting, one can use a pre-trained deep neural network to extract features from images belonging to a scenario different from the one where the pre-training was carried out. In addition, the pre-trained model can be fine-tuned for the new scenario, by using just a few samples available. In fact starting from a pre-trained network, it is possible to achieve high performance in a new (yet similar) scenario with just a few annotated training samples. This procedure moves toward techniques of domain adaptation and few-shot learning described below.

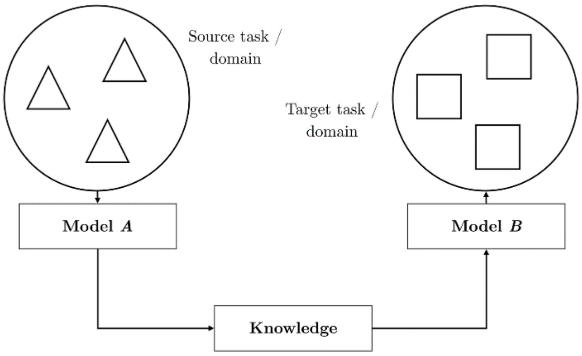


Figure 46: The process of transfer learning[143].

Domain adaptation: Most deep learning based methods require a large amount of labelled data and make a common assumption: the training and testing data are drawn from the same distribution. The direct transfer of the learned features between different domains does not work very well because the distributions are different. Thus, a model trained on one domain, named *source*, usually experiences a drastic drop in performance when applied on another domain, named *target*. This problem is commonly referred to as *Domain Shift*[144] (Figure 47). Domain Adaptation is a common technique to address this problem. It adapts a trained neural network by fine-tuning it with a new set of labelled data belonging to the new distribution. However, in many real cases, gathering another collection of labelled data is expensive as well, especially for tasks that imply per pixel annotations, like semantic or instance segmentation. In this respect, solutions of *Unsupervised Domain Adaptation* (UDA)[145,146] can be used, that do not use labelled data from the target domain and rely only on supervision in the source domain. Specifically, UDA takes a source labelled dataset and a target unlabelled one. The challenge here is to automatically infer some knowledge from the target data to reduce the gap between the two domains.

---

[143] Image source: S. Ruder, The State of Transfer Learning in NLP (2019) https://ruder.io/state-of-transfer-learning-in-nlp/index.html

[144] Torralba, A. and Efros, A. (2011). Unbiased look at dataset bias. InCVPR 2011, pp. 1521–1528

[145] Ganin, Y. and Lempitsky, V. (2015). Unsupervised domain adaptation by backpropagation. Proceedings of Machine Learning Research, pp. 1180–1189, 2015.

[146] L. Ciampi, C. Santiago, J.P. Costeira, C. Gennaro, G. Amato, Domain Adaptation for Traffic Density Estimation. VISIGRAPP (5: VISAPP), 185-195, 2021

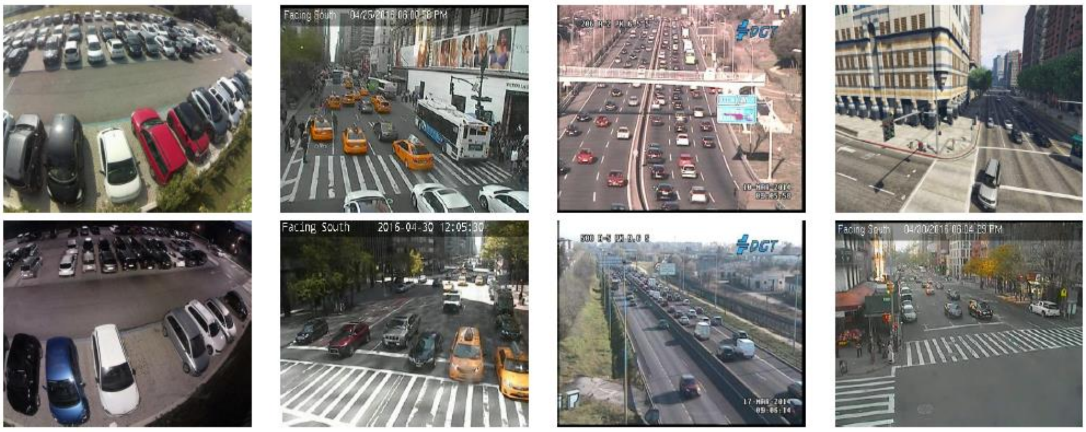


Figure 47: Examples of domain shifts. (a) Day to Night, (b,c) Different points of view, (d) Synthetic to real data[147].

Few-shot learning: Few-shot learning denotes deep learning approaches which are *explicitly* designed to learn from only a few samples per class, starting from a pre-trained model (trained on *base classes*)[148]. Typically, one to ten samples per class are provided for the *novel classes* for which training data is scarce. Few-shot learning methods can be applied to different tasks, like image classification, object detection or semantic segmentation. The existing approaches can be categorised roughly into data augmentation methods, metric learning methods and meta-learning methods. Data augmentation / hallucination methods generate more samples from the few existing ones, e.g. via image synthesis with GANs (Generative Adversarial Networks). Metric learning / embedding methods work by embedding the sample (feature) into a metric space and doing the classification in this space. Meta-learning / optimisation methods aim to pre-train a learner (classifier) so that it can be quickly transformed to the new task setting (e.g., different classes, different domain) in few training steps, enabling it to classify the novel classes.

Self-supervised Learning: Self-supervised learning (SSL) aims at learning from unlabelled data, in a way which is similar to how most of the knowledge learned by humans is believed to be acquired. The idea is to perform a pre-training phase of the network parameters by resorting to an auxiliary task. Examples of these tasks may include trying to reconstruct randomly masked patches from real images[149], trying to identify randomly masked terms from natural language sentences or deciding whether two sentences are consecutive or not[150], to name a few. A

[147] Source image from: L. Ciampi, C. Santiago, J.P. Costeira, C. Gennaro, G. Amato, Domain Adaptation for Traffic Density Estimation. VISIGRAPP (5: VISAPP), 2021, 185-195

[148] Yaqing Wang, Quanming Yao, James Kwok, Lionel M. Ni, Generalizing from a Few Examples: A Survey on Few-Shot Learning, arXiv:1904.05046v3 , 2020

[149] Hinton, G. E., Osindero, S., and Teh, Y.-W. A fast learning algorithm for deep belief nets. Neural computation, 18(7):1527– 1554, 2006.

[150] Devlin, Jacob, et al. "Bert: Pre-training of deep bidirectional transformers for language understanding." arXiv preprint arXiv:1810.04805 (2018).

notorious case of SSL is called "*Contrastive SSL*" [151,152],and consists of solving an auxiliary task defined upon positive and negative pair examples which are generated in an unsupervised way around the concepts of "same" and "different". The method consists of producing variations of data items (e.g., in the realm of computer vision, by applying rotations, colour or tone variations, etc.) that are assumed to preserve the original class label (e.g., an image of a rotated bird is still to be classified as a bird) despite the fact that the label itself is unknown (e.g., even if the image is not tagged as containing a bird). The problem is then translated to learn to distinguish between the "same" and "different" concepts, by minimising the distance between positive pairs and maximising the distance between negative pairs (the contrastive loss). While the self-supervised phase is typically carried out using as many unlabelled data as possible, the resulting network is ultimately fine-tuned using the (likely few) annotated data at one's disposal.

Synthetic data generation: Although a large amount of annotated data is already available and successfully used to produce important academic results and commercially viable products, there is still a huge number of scenarios where laborious human intervention is required to produce high-quality training sets. To address this problem and make up for the lack of annotated examples, the research community has begun to increasingly leverage the use of programmable virtual scenarios to generate synthetic visual data sets as well as associated annotations. For example, in image-based deep learning techniques, the use of a modern rendering engine (i.e. capable of producing photo-realistic images) has proven to be a valuable tool for the automatic generation of large data sets[153] (Figure 48). The advantages of this approach are remarkable. In addition to making up for the lack of data sets in some particular application domains, these synthetic datasets do not create problems with existing laws about the privacy of individuals related to facial detection, such as the General Data Protection Regulation (GDPR). These techniques have been successfully used in scenarios that include self-driving cars, the detection of safety equipment, the detection of fire arms, etc. They can also be of great help in the media sector to quickly develop solutions able to analyse data to classify and recognise events, objects, actions, which were not considered before and for which no training sets are readily available.

[151] Jaiswal, Ashish, et al. "A survey on contrastive self-supervised learning." Technologies 9.1 (2021): 2.
[152] Chen, Ting, et al. "A simple framework for contrastive learning of visual representations." International conference on machine learning. PMLR, 2020.
[153] M. Di Benedetto, F. Carrara, E. Meloni, G. Amato, F. Falchi, C. Gennaro, Learning accurate personal protective equipment detection from virtual worlds, Multimedia Tools and Applications 80 (15), 23241-23253

Figure 48: Examples of synthetically generated images that can be used for training AI models[154].

Research challenges

Research during the last decades has provided significant advancements in AI. However learning in scenarios with scarce data, still represents a serious challenge and an opportunity for further research directions.

Nowadays, plenty of pre-trained models specialised to work with specific media and specific applicative scenarios are available. One difficulty, typically encountered, is the choice of the best pre-trained model to be used as a starting point to execute transfer learning or domain adaptation. Different models can lead to different performance in different application domains and in some cases several trial-and-error attempts have to be executed to achieve satisfactory results.

In addition, a challenging research direction might be transferring knowledge inferred from a media, to a different one. For instance transferring knowledge acquired in image analysis to analyse different media, such as 3D models, or even text. Cross-modal solutions are being used to allow searching for images using text, or vice versa. Can we extend these methods to cross-modal transfer learning? For instance, is it possible to learn how to extract knowledge from images, once we know how to extract knowledge from text?

In real applications, we also observe a gap between the common practice used to set up benchmarks for research purposes and the real scenarios. For instance, in case of few-shot learning techniques, we can observe three main aspects, where the setup of benchmarking problems (and thus the methods described in literature, as well as the existing implementations available from the research community) deviate from the practical requirements of using few-shot learning in media use cases:

- The typical setup of the problem is posed as *n*-way *k*-shot, i.e. a problem with *n* classes and *k* samples per shot. However, in practice the number of samples per class that are provided may differ.

[154] Image source: M. Di Benedetto, F. Carrara, E. Meloni, G. Amato, F. Falchi, C. Gennaro, Learning accurate personal protective equipment detection from virtual worlds, Multimedia Tools and Applications 80 (15), 23241-23253

- There is not a fixed predefined dataset, but the set for base classes will contain a mixture of third party and maybe own proprietary data for some classes, while the novel classes are mined from own or third party media content (e.g., web sites). Thus, the concept of a dataset is fluid and the available data will evolve over time.
- Classes need to be added incrementally, which requires creating balanced training sets, but approaches should aim to keep the training effort low. This again means that there is no fixed notion of a dataset, but the dataset needs to be updated on the fly via some sort of incremental learning.

Societal and media industry drivers

Vignette: Tagging and learning new classes for audio-casual content search

Valerie and Theodor are working in the archive of a large broadcaster. Valerie is responsible for supporting the journalists and editors in the newsroom of the broadcaster looking for content, whereas Theodor documents content arriving in the archive with automatically and manually generated metadata. Based on the incoming queries from the newsroom for material of a certain kind (e.g. showing a certain person / location / object / entity), Valerie does a search in the broadcaster's archive for proper material and returns matching video clips to the newsroom editors. Due to the large size of the video archive (> 1 million hours), an exhaustive content-based search cannot be done in the provided timeslot. Therefore, usually the search in the archive is done by searching for certain (textual) metadata tags derived from the associated query. For a successful search, this means that it is crucial that new content which is added to the video collection is annotated properly (documented with the proper concepts). Of course, a concise manual annotation of a video is a very time-consuming process, therefore Theodor employs a semi-automatic workflow with the help of an AI tagging engine. The AI tagging engine employs automatic video analysis methods (object detection, face recognition etc.) internally and proposes a list of metadata tags based on the result of the video analysis. These proposed metadata tags are now inspected by Theodor and either accepted or discarded by him.

In February 2020, a new virus named SARS-CoV-2 has appeared suddenly and is spreading very fast throughout the world. Therefore, nearly every day the broadcaster is reporting in its news program about this new virus, which has the potential for a pandemic. Consequently, also Valerie is getting queries for content related to SARS-CoV-2 every day from the newsroom. Besides other related content (such as queries for street shots with people wearing face masks), the editors in the newsroom are also interested in video content which is showing visualisations of the SARS-CoV-2 virus with the characteristic spikes on its surface. As this kind of object is not existing yet in the AI tagging engine, it is not able yet to automatically detect and tag the object "SARS-CoV-2 virus" in a video. Which means more work for Theodor, as he has to tag the new object class manually. But the AI tagging engine provides a comfortable mechanism for incrementally adding new object classes, via few-shot learning. For this, all Theodor has to do is to provide a few (five to ten) images of the desired object class that he has annotated in recent days. After that, the new object class is incrementally trained in very short time (less than 15 minutes), and the updated AI tagging engine is now able to detect and automatically tag also the object class “SARS-CoV-2 virus shape” in video content. He will now tag a few of those face masks, to also support it in the object detector, and being able to discriminate between street scenes where people wear masks or not. As face masks are more challenging to identify visually, more samples are needed to increase the robustness, but those can be obtained from tracking faces throughout the video, and thus obtain a variety of poses of the masks.

Future trends for the media sector

In the media sector, technological solutions should constantly be aligned to the evolution of global affairs. Consider, for instance, news production: Media companies, in this case, should immediately intercept new emerging trends, new hot topics, and new relevant events. Accordingly, an AI system, analyzing the continuous flow of multimedia digital information constantly generated, cannot rely on pre-determined tools of analysis, but they have to be constantly updated and trained with the scarce and noisy data at hand.

As an intuitive example, imagine a face recognition tool that has to intercept pictures of a person who just recently became popular for some reason and who was unknown until yesterday. A pre-trained model of a face recogniser would not be able to recognise unknown people. Therefore, analysis tools (classifiers, face recognisers, etc.) should be immediately updated to be able to react quickly to this. Similar examples can be done, considering text, image, and video classifiers, that are supposed to identify occurrences of media related to new emerging topics or to detect fakes and disinformation in social media.

Effective solutions able to learn with scarce data allow media companies to immediately react to the occurrence of new emerging topics and be able to effectively detect media related to them, with minimum effort, minimum delay, and high accuracy.

Goals for next 10 or 20 years

In the future, we will need AI tools that are able to learn from the user with as little effort as possible. Nowadays, most of the effort needed to produce accurate AI tools relies in the preparation of a high-quality training set. In the future, solutions that learn from noisy and evolving data, in a trustable and reliable way, might represent the real goal. Solutions based on reinforcement learning and learning with scarce data go in this direction and can provide a significant step forward in AI.

Humans learn from their daily experiences and from interactions with other humans. Once we become experienced in a field, we require little effort and training to adapt to new related tasks. In addition, working in a team can further improve our skills.

We can imagine adopting a similar paradigm also in AI systems in the next decade. We can imagine an ecosystem of AI tools that interact and continuously improve their performance by autonomously learning from other tools and from available (possibly scarce) data, by constantly staying up to date with the most recent and emerging skills they need for their assigned tasks.

5.4 Transformers for computer vision tasks

Contributors: *Ioannis Pitas (AUTH), Vasileios Mygdalis (AUTH)*

Current status

During the last decade, we have experienced the vast power of different types of deep neural networks in a variety of application tasks. Transformer models constitute a novel type of artificial neural networks that joined the family of already well-established neural network architectures such as multi-layer perceptrons (MLPs), convolutional neural networks (CNNs), recurrent neural networks (RNNs) and many more. ***Transformers***[155], originally introduced for language machine translation, demonstrated unprecedented performances in a range of NLP tasks such as text classification, text summarisation and question/answering systems. One such famous language model is the GPT model from OpenAI[156], which is capable of generating text that is rather difficult to determine whether or not it is written by a human. Following the breakthroughs in the NLP domain, a great interest emerged from the computer vision community in order to adapt Transformer models in computer vision. To this extent, during the last two years, Transformers have been applied in some standard computer vision tasks achieving superior results.

So far, CNNs have been the de-facto approach for tackling computer vision tasks. However, the convolution operation comes with certain shortcomings, including their inefficacy to capture long-range dependencies such as relations between pixels in an image that are distant. For example, an early convolution layer of a model trained to recognise faces can encode information about whether certain face features such as “eyes”, “mouth” or “nose” are present in an image, but these representations will not contain information such as “eyes are above nose” or “mouth are below nose”. The formidable power of Transformers, on the other hand, derives from the ***attention operation***. Inspired by the concept that humans tend to pay attention to certain factors when processing information, the attention mechanism was developed in order to direct neural networks to focus only on important parts of the input data (Figure 49). The attention mechanism utilises an entire image as context, as opposed to the convolution operation which is meant to operate on a small fix-sized region at a time. Moreover, the attention mechanism enhances the most important parts of an image while simultaneously discarding the rest. Despite their apparent higher complexity, Transformer models with attention have yielded state-of-the-art results in some standard computer vision tasks such as image recognition, object detection, image segmentation, image generation and much more.

[155] A. Vaswani et al., “Attention is all you need,” in Advances in neural information processing systems, pp. 5998–6008, 2017.
[156] T. B. Brown et al., “Language models are few-shot learners,” arXiv preprint arXiv:2005.14165, 2020.

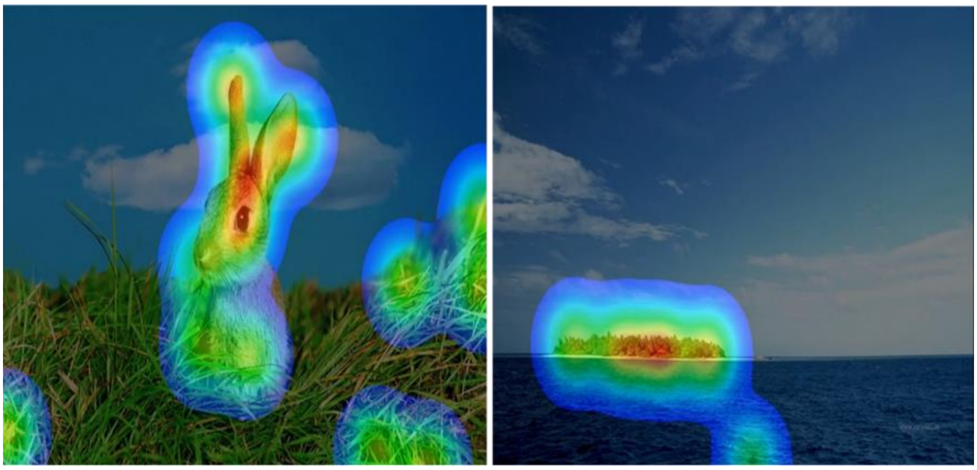


Figure 49: Visualisation of the attention operation applied on images [157].

Research challenges

Despite their excellent performance in the aforementioned visual tasks, Transformer models suffer from certain challenges and limitations associated with their applicability to practical settings. Perhaps their most prominent bottlenecks include a) the requirement of a vast amount of data (either labelled or unlabelled) for training purposes, and b) the associated high computational complexities derived from the attention operation.

As far as the first bottleneck is concerned, it has been shown that, depending on the visual task at hand a Transformer model requires from a few hundred thousand to hundreds of millions training images. For example, the ViT model[158] requires over 300 million of image examples in order to achieve comparable performance on the ImageNet benchmark dataset. It is obvious that for many specific applications (e.g., the recognition or detection of a rare animal species) it is not possible to obtain the required amount of training images. In such a scenario a common trick is to utilise information from a pre-trained model that has been trained with adequate training samples in a different domain (see section 5.3).

The second bottleneck is strictly related to the high time and memory costs imposed by the attention operation. As already discussed, attention utilises the entire images in order to extract meaningful information regarding their relations. Furthermore, it is obvious that when dealing with images the working dimensional space can grow exponentially and so do the aforementioned costs. This can lead to high training and inference times that in certain cases become unacceptable. However, the questions of designing efficient, low-complexity

[157] Image taken from: X. Hou et al., "Saliency detection: a spectral residual approach," Proc. IEEE CVPR, p. 4270292, June 2007.
[158] L. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020.

Transformers that can moreover work in a data-efficient manner are open research problems and recent works report encouraging steps towards their resolution.

Societal and media industry drivers

Vignette: Visual content retrieval/manipulation in large video archives & productivity enhancement

Albert is working for a televised news broadcaster and is responsible for creating short video clips/segments that will thereby be shown during the typical evening news report. It is Friday noon, and there was a devastating flood in Thailand that took place just a few hours ago that demolished 50 buildings. His manager Mary asked him to edit a short 45-second video clip to cover the story for today's evening live broadcast. At the time, Albert had access only to a 10-second short video clip obtained by a mobile phone at the scene, and he should enrich it by incorporating content from a wealth of previously covered clips covering the theme "flood" from all over the world, stored in the broadcaster archives. Unfortunately for him though, the archive was not well documented thus he did not know which of the archived video material included floods with building demolitions. To retrieve similar video content, he runs a Transformer-based computer vision model where he employs the short clip of the mobile phone as a query. During his lunch break, the system ran through the broadcaster's rich archive and retrieved the most visually similar video segments. The retrieval was accurate because it recognised semantic information from the mobile phone clip (i.e., building, water) and matched it with the one from the stored material. However, the archived video segments were in a different resolution compared to the one captured by the mobile phone. This is not a problem for Albert though, because another model smooths out visual inconsistencies and creates a visually pleasing and coherent video result, in terms of automatic brightness adjustment, colour restoration and image upsampling. Within a short time-frame, Albert carefully stitches the video segments, adds some visual effects and his work is effectively done.

Future trends for the media sector

It becomes apparent that Transformer networks are all geared up to dominate the world of computer vision as more and more institutes and companies are attracting their attention. Since their introduction for simple computer vision tasks, such as image recognition, Transformers have been applied to some more advanced and complex tasks such as object detection[159], object tracking[160], image/instance segmentation[161] and depth estimation[162]. All of the aforementioned tasks can be applied to a wide range of real-life applications, from agriculture and road

[159] N. Carion et al., "End-to-end object detection with transformers," arXiv preprint arXiv:2005.12872, 2020.
[160] T. Meinhardt et al., "TrackFormer: Multi-Object Tracking with Transformers", arXiv preprint arXiv:2101.02702, 2021.
[161] L. Ye et al., "Cross-modal self-attention network for referring image segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019.
[162] L. Zhaoshuo, et al. "Revisiting stereo depth estimation from a sequence-to-sequence perspective with transformers." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.

inspection to traffic surveillance and search and rescue missions. The complete or even partial automation of such applications enabled by Transformers can eliminate or minimise any associated logistic costs, complicated requirements and personnel involved.

Another important computer vision field which Transformer-based neural networks have excelled is that of image generation[163]. This field consists of several interesting subfields such as image style transfer, image colourisation, image reconstructions, image super-resolution, and image synthesis (Figure 50). It has been demonstrated that Transformer models are able to generate coherent and realistic images or even artistic-like ones that can subsequently assist in the creation of a whole range of media content such as graphics for computer games or even real-action and animated movies and news videos. This increased interest in visual content generation is also amplified by the ever-expanding use of image/video-based social media platforms such as Instagram and TikTok where users are always excited and thrilled for the creation of new content.

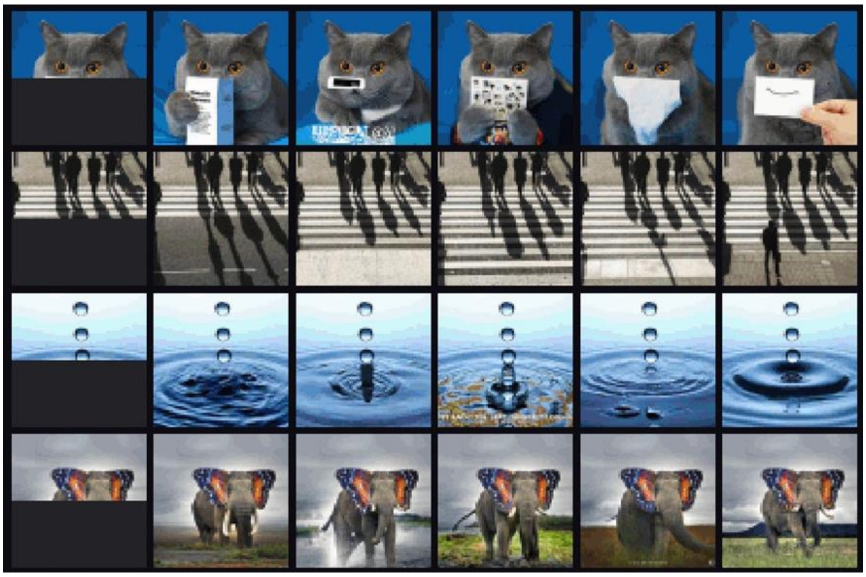


Figure 50: Image reconstruction with Transformer model[164].

Goals for next 10 or 20 years

Right now, Transformer neural networks represent one of the field's most promising present-day advances. The research on this novel type of deep neural networks is attracting more and more attention both from the academia and the industry and this attention is expected to rise significantly as, per many leading pioneers, we have only just scratched the surface. Inspired by their successes in a wide range of different individual downstream tasks, such as modelling

[163] Y. Jiang et al., "TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up", arXiv preprint arXiv:2102.07074, 2021

[164] Image taken from: M. Chen, et al. "Generative pretraining from pixels." International Conference on Machine Learning. PMLR, 2020

natural language, images, proteins and behaviour amongst few the trend has shifted towards exploring ways to constitute Transformers as universal computation engines so that a single trained model can generalise and solve a number of multiple tasks from different data domains. Transformer networks are enabling a variety of new applications and their continuous rise is a critical and important factor in the advancement of AI in the future years.

More specifically, in the next 5 years, we expect that Transformer-based architectures will become the norm in semantic metadata extraction and content retrieval, having important applications in the media industry for video archiving and indexing. In addition, within the next 10 years, due to their ability to model spatio-temporal and multi-modal relationships in a structured manner, they will become very useful in audiovisual data summarisation and will foster the automation of workflows in audiovisual content creation.

5.5 Causality and machine learning

***Contributors**: Filareti Tsalakanidou (CERTH)*

Current status

During the last two decades, machine learning algorithms have had enormous success in prediction tasks that require high-dimensional inputs, including computer vision and natural language processing. This success can be attributed to "large-scale pattern recognition on suitably collected independent and identically distributed data"[165]. However, in the real world there's very little control on data distributions, which may cause for example object detection algorithms to fail in the presence of illumination and pose variations or camera noise, issues that humans can overcome with no or very limited effort due to their inherent ability to generalise and also transfer their knowledge from one setting or domain to another.

The ability to generalise or adapt is only one of the current shortcomings of machine learning. Another limitation is explainability. Currently, ML models are mostly black boxes that produce predictions or recommendations without usually being able to provide a clear explanation as to why they predict A instead of B. This may not be considered much of a problem when an online bookstore recommends a science fiction book instead of a history book but it becomes vital when medical, economic, social or environmental predictions or recommendations are made that have real tangible impact on people's lives. Attempts for interpretation mainly focus on how the ML model works, not how the world works thus failing to give an answer to the why question. Why one predicted outcome is more possible than another? Why one recommendation may be better for a specific person?

[165] B. Schölkopf et al., "Toward Causal Representation Learning," in Proceedings of the IEEE, vol. 109, no. 5, pp. 612-634, May 2021, doi: 10.1109/JPROC.2021.3058954.

Moreover, what current AI systems are missing is the ability to understand ***cause-and-effect (causal) relationships*** between actions and to consider ***counterfactuals***, i.e. engage in what-if questions about possible past and future chains of events that would lead to different outcomes. This fundamental element of human intelligence, and a key to human evolution, is still missing from AI systems, limiting their ability for independent and autonomous decision-making.

Causality (or cause and effect) is the "relation between two events, one of which is the consequence (or effect) of the other (cause)"[166]. Cognition of causality is fundamental to human development. Humans start to understand such cause-effect relations from infancy, by observing the world around them and learning to predict/expect the consequences of different actions and events. Research has shown that as young as 6 months old, infants are able to "categorically perceive motion events along causal dimensions in addition to spatial and temporal dimensions"[167].

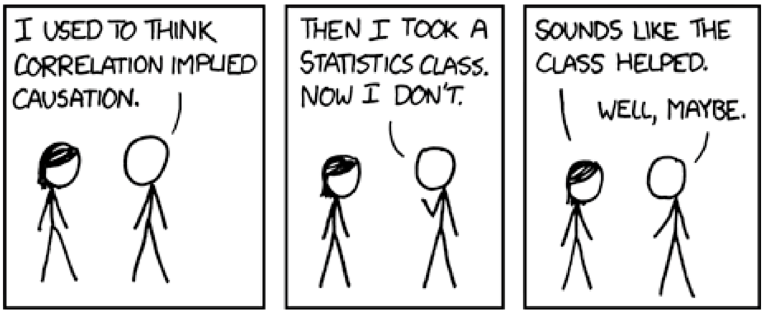


Figure 51: Correlation and causation (comic by Randall Munroe[168]).

Unlike humans, however, machine learning algorithms struggle with causality, failing to understand and determine even basic causal inferences. Take a video of Rafael Nadal and Roger Federer playing tennis as an example[169]. A human can understand that when Nadal's racket clashes with the tennis ball it will cause the ball to change its direction and go back to Federer's side. They are also able to consider counterfactuals, e.g. what would happen if Federer's shot was too low or too high for Nadal to hit the ball and how would that affect the game. Trained with millions of examples, ML algorithms can segment thousands of such video frames in seconds, detect and label different objects like humans, racquets, balls and nets in real-time, provide accurate video summaries or audio-to-text transcriptions but have difficulties in understanding basic causal relationships that a toddler can intuitively comprehend. As explained above, this difficulty is inherited in ML approaches and is a result of the main assumptions and learning tactics adopted by such methods.

[166] Bender, A. (2020) What Is Causal Cognition? *Frontiers in Psychology* 11(3). doi: 10.3389/fpsyg.2020.00003
[167] Muentener, P., and Bonawitz, E. B. (2017). "The development of causal reasoning" in The Oxford handbook of causal reasoning. ed. M. R. Waldmann (New York: Oxford University Press), 677–698.
[168] Illustration from XKCD available at https://xkcd.com/552
[169] Example inspired by B. Dickson, Why machine learning struggles with causality (2021): https://bdtechtalks.com/2021/03/15/machine-learning-causality/

Traditional ML approaches are excellent in identifying patterns and associations (correlations in the collected data) and predicting outcomes given a sufficient amount of data. However, to really take AI research to the next level and address the limitations of ML, we need to combine such traditional techniques with causal inference methods so as to be able to move beyond simple correlations to identification of causes thus, being able to answer why something happens and also transfer the acquired knowledge about causes and effects to other relevant domains (Figure 51).

Research challenges

Causality is discussed in J. Pearl's "*The Book of Why: The New Science of Cause and Effect*"[170], which argues about the need to move beyond existing data-centric ML approaches and endow AI with causation capabilities. Pearl proposes the ***Ladder of Causation***, which constitutes a three-step approach to achieve real AI (Figure 52). The first step is "***seeing***"; here AI can learn from data and ***find associations*** that allows it to make accurate predictions. For example, the editorial board of a newspaper would like to predict which of the candidate guest opinion articles will produce the most reader engagement based on past data of user engagement with their website and social media. This is where the majority of ML algorithms currently operate, offering statistical predictions that are only accurate if the conditions remain the same. In our example, a recommender trained with past article engagement data would have failed to offer an accurate prediction about the interestingness of a medical article on viruses when the pandemic had just begun.

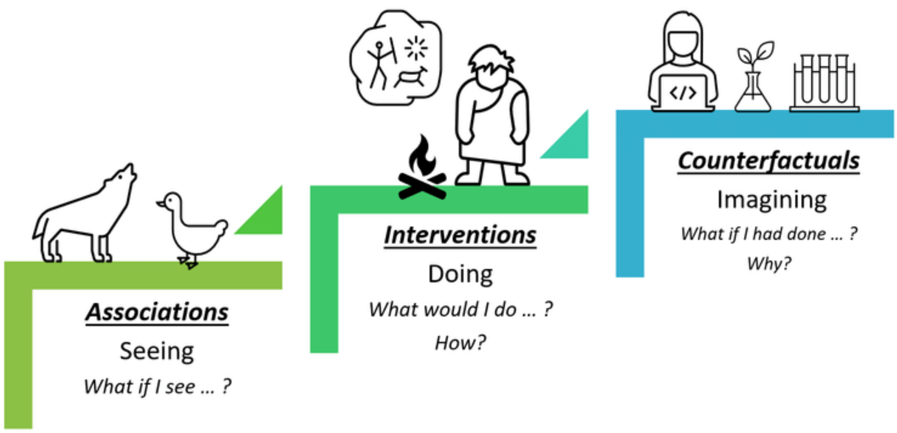


Figure 52: The Ladder of Causation[171]

Second step is "***doing***", which involves ***interventional reasoning***, i.e. trying to predict the outcomes of specific actions (interventions) that change the current conditions. For example,

[170] Judea Pearl and Dana Mackenzie. 2018. The Book of Why: The New Science of Cause and Effect (1st. ed.). Basic Books, Inc., USA.

[171] Image taken form Carey, Alycia & Wu, Xintao. (2022). The Fairness Field Guide: Perspectives from Social and Formal Sciences (source file: https://www.researchgate.net/figure/Pearls-Ladder-of-Causation-The-first-rung-associations-only-allows-predictions-based_fig2_357875366)

the editorial board considers publishing more articles of younger guest columnists inspired by music and entertainment industry trends – how will this decision affect readership among older audiences; will it attract younger audiences? In terms of machine learning, we are dealing with a distribution shift that changes the underlying conditions and thus the statistical relations between variables. Obviously, this type of prediction requires the development of a causal model that understands the causal relations that affect reader engagement with an article, considering many different variables.

The third step is "***imagining***" and involves ***counterfactual reasoning***, i.e. examining different scenarios of what could have happened if the conditions were different. For example, the editorial board would like to know what the effect on readership would have been, had they not published so many articles critiquing the government before the last election. These kinds of predictions require causal models that depend also on unobserved data. They are critical in AI since they allow to reflect on past decisions and test hypotheses.

Causal AI is still in a nascent state but the field is expected to grow significantly in the next years, aiming to overcome existing ML limitations and deliver more human-like machine intelligence. Unlike machine learning models, causal models describe the causal mechanism of a system based on observed data, a set of variables and assumptions. They are able to incorporate data distribution changes when interventions are applied to the system thus being better in generalising than ML models[172]. The main approaches proposed include Causal Bayesian Networks (CBN) and Structural Equation Modelling (SEM)[173]. These causal models comprise of a statistical model and a causal graph that reveals the causal relations between the different variables. To identify the causal relations, two types of search algorithms are proposed[174]: the first exploits conditional independence relations in the data to find a Markov equivalence class of directed causal structures; the second finds a unique causal structure under certain assumptions, considering a noise term that is independent from causes. Newest methods try to exploit deep learning techniques to recognise simple cause-and-effect relationships[175]. Recently, Netflix proposed computational causal inference (CompCI) as an interdisciplinary field across causal inference, algorithms design, and numerical computing that "*addresses engineering needs and human needs for scalability, and directly benefits the deepening relationship between*

---

[172] J. Ramachandran, Causal Learning – The Next Frontier in the Advancement of AI (2021): https://www.course5i.com/blogs/causal-learning-in-ai/

[173] Peter Spirtes (2010). Introduction to Causal Inference. *Journal of Machine Learning Research* 11(54), 1643–1662.

[174] Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524

[175] Y.Bengio, T. Deleu, N. Rahaman, R. Ke, S. Lachapelle, O. Bilaniuk, A. Goyal, C. Pal (2020). A meta-transfer objective for learning to disentangle causal mechanisms in Int. Conf. on Learning Representations 2020 (ICLR2020), https://openreview.net/pdf?id=ryxWIgBFPS

experimentation and personalisation in products and algorithms"[176]. Recent surveys on causal discovery[177,178] provide a good overview on both background theory and proposed methods.

Identifying causality in real-world settings by exploiting the recent breakthroughs in deep learning and data science is the main challenge of the field. This will hopefully allow efficient understanding of the underlying relations of different variables and thus of how a real-world system, physical or other, operates as a whole. The integration of causal reasoning and inference has the potential to improve the generalisation capabilities of AI and facilitate easy domain adaptation, exploiting the identified causal relations between variables. For example, in the game industry agents currently have to learn to play a game from scratch, even if the game is of the same genre and includes similar strategies. Causal AI can help distill previous knowledge in new games, similarly to a young gamer that uses her previous playing experience but also other life experiences to play a new game. An additional challenge that causal AI inspires to address is explainability of current ML methods, aiming to demystify and provide clear explanations on how decisions are made and actions are driven in AI systems that currently seem like a black box to the average user and in many cases to the experts that have developed them.

Fairness and bias mitigation is another challenge of AI systems where causality can help. According to J. Loftus[179], fairness and causal inference are dual problems since in causal inference, you try to understand the effect of a certain variable while in fairness you try to make a certain variable (like gender or race) not have any effect. To this end, counterfactual fairness in models can be ensured by using causal models.[180] As M. Kusner points out[179], causal models will allow "*to reimagine any individual as a different race or gender, or any different attribute, and make a prediction on that imagined person*." Ensuring that algorithmic outcomes are the same in the real world and a 'counterfactual world' will lead to fairer AI.

According to J. Bengio, one of the pioneers of deep learning, after the great success of ML-based AI what we need now are "*machines that understand the world, that build good world models, that understand cause and effect, and can act in the world to acquire knowledge*"[181,182].

Societal and media industry drivers

Vignette 1: Making successful programming choices in big TV networks

[176] Jeffrey C. Wong (2020). "Computational Causal Inference", Netflix: https://arxiv.org/abs/2007.10979
[177] Glymour, C., Zhang, K., & Spirtes, P. (2019). Review of Causal Discovery Methods Based on Graphical Models. *Frontiers in genetics*, *10*, 524. https://doi.org/10.3389/fgene.2019.00524
[178] Matthew J. Vowels, Necati Cihan Camgoz and Richard Bowden (2021), D'ya like DAGs? A Survey on Structure Learning and Causal Discovery, https://arxiv.org/abs/2103.02582v2
[179] The Alan Turing Institute, Fairer algorithm-led decisions (2018): https://www.turing.ac.uk/research/impact-stories/fairer-algorithm-led-decisions
[180] M. Kusner, J. Loftus, C. Russell, R. Silva, Counterfactual fairness (2018) at https://arxiv.org/abs/1703.06856
[181] B. Dickson, System 2 deep learning: The next step toward artificial general intelligence (2019): https://bdtechtalks.com/2019/12/23/yoshua-bengio-neurips-2019-deep-learning/
[182] YouTube video "Yoshua Bengio: From System 1 Deep Learning to System 2 Deep Learning (NeurIPS 2019)": https://www.youtube.com/watch?v=T3sxeTgT4qc&ab_channel=LexClips

Emilia is the Head of the Entertainment section of a TV network and she has to make a lot of different decisions on a daily basis: select the pilots they will order to series, order new pilots based on submitted scenarios, make decisions about casting of new shows, adjust programming to increase ratings, select which shows will be offered in the network's streaming platform, decide which shows need further promotion and what kind of promotion, envisage ways to beat the competition, etc. Emilia and her team are assisted by an AI TV programming assistant with causal capabilities. The assistant has been trained with heterogeneous historic user engagement data (TV ratings, ratings on sites like Rotten Tomatoes and IMDB, online reviews by TV critics and audience, social media comments, screenings with live audience, surveys, etc.) for the network but also other networks, streaming services, production companies and content creators/ distributors, etc. and integrates causal models from different domains (like psychology, social sciences, advertisement, etc.). Given that the ratings of the prime-time zone on Thursday evenings have been dropping recently, Emilia needs to make quick changes to the schedule. To this end, she asks the AI assistant to determine the possible causes for this drop. The assistant identifies two main causes: the legal drama at the 9:00-10:00 pm slot seems to have caused a lot of online critique for promoting an image of the legal and incarceration system that contradicts the living experiences of many people in minority communities while also failing to embrace diversity in terms of casting. In addition, in a competitor network a new sci-fi series seems to draw the younger audience to that channel.

Emilia is faced with different questions. To beat the competition, should she re-schedule the legal drama to another night and put in its place a high-ratings adventure series with significant following among young audiences that currently plays on Mondays? Will that cause the ratings of the adventure series to temporarily fall? How will it affect advertisers? Or should she urgently put on the Thursday time-slot one of the new shows that were scheduled for later this year? Which one of these shows has the potential to beat the competitors' sci-fi show? Why is the audience drawn to that show? With regard to the critique for the legal drama, would different storylines be enough? Or should they also introduce new cast members? The AI assistant is able to examine these scenarios and estimate the outcome on ratings based on a series of intervention and counterfactual scenarios. Based on the analysis, Emilia decides to promote one of the new shows in the Thursday timeslot, taking her chances with a romantic comedy with sci-fi elements and a diverse cast that seems to appeal to a young female audience. In addition, she decides to give the green light for changes in the cast of the legal drama based on the suggestions offered by the AI assistant about the profile of new cast members and a relevant analysis of how different elements of diversity affect ratings on different audiences.

Vignette 2: Improving newspaper readership and newsroom operation using causal AI for successful editorial choices

Ariadne is the new Editor in Chief of a big newspaper. She was hired because both readership and online engagement of users had dropped considerably while trust in the medium as estimated by a relevant survey had fallen significantly. Ariadne must analyse the reasons why this happened and take measures to reverse the situation. She is assisted by an AI newsroom assistant that provides a full analysis on the main reasons readership has fallen. The report includes interesting findings: the audience has stopped reading the articles of specific political commentators because they consider them partisan; articles about the pandemic attract fewer readers because readers are tired of the newspaper's highly-negative coverage; younger audiences consider the newspaper 'old' because of lack of diversity and different voices; online audiences are not satisfied by the format of the online experience, etc. Ariadne decides that a revamp is in order but first wants to understand how journalistic and editorial decisions were made up to now. The newsroom assistant offers a causal analysis of the factors that influence the journalistic choices of different editors and journalists in the newsroom, offering evidence of influence by political interests and of biases with regard to gender and age. Ariadne considers the evidence provided and makes some first decisions to re-assign journalists to different topics and enforce stricter rules with regard to non-partisan coverage of political news. In addition, she considers hiring a new group of diverse journalists to cover social issues but also entertainment and media, aiming to engage younger readers and adopt a less political profile for the newspaper. Ariadne also hires a graphical designer to redesign the newspaper website aiming to maximise appeal on younger but also older audiences. Using causal inference, the AI newsroom assistant as well as an AI user experience assistant (an AI system that uses causal inference to estimate the impact of platform design changes to user experience and test counterfactual scenarios) estimate how these interventions may affect readership and engagement among different demographics as well as advertisement. After a few months, the decisions made by Ariadne are re-evaluated and a set of counterfactual questions is examined by the AI assistants to estimate what could have been different if different decisions were made.

Future trends for the media sector

Causality can be a game changer in the media sector in several ways. We highlight some of these opportunities with an eye to AI4Media use cases:

- ***Improve personalisation of services*** and ***increase user engagement*** by capturing the causal relations between services, offered content, user interfaces, user profile/ behaviour, global/local trends, advertisement strategies, etc.
- ***Enhance recommender systems*** by a) identifying causal relationships between user attributes, user choices, and media content to estimate the user reaction when exposed to different content, b) mitigating biases through counterfactual testing, c) explaining why specific content was recommended.
- Help the fight against disinformation by ***understanding the mechanisms for disinformation spread and influence on social media*** (including disinformation topic, susceptibility of different groups of people, role of networks of friends, role of politician/political ideology/polarisation, impact of social media and traditional media, format of disinformation content, effect of fact-checking etc.) exploiting existing

knowledge from psychology and social sciences[183,184]. In addition, it can improve the trustworthiness of existing deepfake or disinformation detection tools by providing explanations on why some content is considered false or fabricated.

- ***Understand how journalistic and editorial decisions are made*** and which are the factors that influence them the most, by considering personal behaviours in the newsroom, media organisation structure and codes of conduct, political/social/economic national and international environment, influence of advertisers or economic and political interests, public opinion trends, cultural differences, systemic biases etc.[185]
- ***Improve general or targeted advertisement*** and customise campaigns for different products or issues by detecting the main drivers that will convince different groups of customers and estimating the impact of different campaign strategies based on these findings.
- Produce ***automatic summaries of movies, sports or user videos*** that will explain not only who or what we see but how the action or inaction of different characters affects the plot or what relationships the different characters form between them.
- ***Solve the generalisation problem*** of current object detection algorithms thus allowing detection of objects in different settings and under different conditions and improving significantly search efficiency in large audiovisual databases.
- ***Improve and facilitate game design*** by a) allowing gaming experience from one game to be transferred to another, b) integrating causal models for real-world social interactions to the gaming environment to improve human-machine interaction and allow for natural story-telling, c) identifying why different groups or types of players interact in different ways or adopt different strategies, aiming to maximise satisfaction and engagement, d) estimate how adoption of different game rules, strategies or capabilities will affect player engagement, e) estimate how different game elements may affect the mental health of users exploiting existing knowledge from psychology.[186]
- ***Remove bias and improve explainability of content moderation systems*** by understanding whether moderation policies are biased against specific groups or beliefs and by providing explanations about why some content is automatically removed. Causal reasoning and counterfactuals can also help refine moderation strategies by estimating the impact of moderation policy changes and assessing past decisions.

Goals for next 10 or 20 years

[183] Lu Cheng, Ruocheng Guo, Kai Shu, Huan Liu, Causal Understanding of Fake News Dissemination on Social Media (2021) at https://arxiv.org/abs/2010.10580

[184] S. T. Smith, E. K. Kao, D. C. Shah, O. Simek and D. B. Rubin, "Influence Estimation on Social Media Networks Using Causal Inference," 2018 IEEE Statistical Signal Processing Workshop (SSP), 2018, pp. 328-332, doi: 10.1109/SSP.2018.8450823

[185] Lukas P. Otto and Isabella Glogger (2020) Expanding the Methodological Toolbox: Factorial Surveys in Journalism Research. *Journalism Studies 21*, 947-965.

[186] Katarina Gyllenbäck, Putting into play – A model of causal cognition on game design (2018): https://katarinagyllenback.com/2018/12/17/putting-into-play/

In the next couple of decades, AI applications will be endowed with causal and inference capabilities, while repositories of causal models will be available, modelling everyday phenomena but also expert concepts thus allowing the development of new AI systems with efficient understanding of causes and effects in new domains. AI systems will be able to discover new causal models and to verify and improve them based on ML and collected evidence from active experimentation in a new media domain.[187,188]

5.6 AI at the edge

***Contributors**: Symeon Papadopoulos (CERTH), Emmanouil Krasanakis (CERTH)*

Current status

Technological improvements of the last decades have led to a widespread adoption of smart devices, such as mobile phones and sensors, in everyday life. For example, there are over 3 billion Android devices[189] that run software catering to a variety of needs, such as web surfing, creation and consumption of multimedia, social networking, and analysis of sensor data that range from weather readings to biometric ones. Many software products make use of AI breakthroughs to enhance user experience, for instance by recommending multimedia content or social interactions and automatically generating short description or tag summaries. These operations are often supported by central services, which are accessed through internet endpoints and store and process user data, for example for the purposes of retrieving those upon request or performing AI inference.

The above-described dependence on central endpoints makes software reliant on third-party infrastructure and services that make users hand over control of their private data. However, in our increasingly digital societies, the needs for data privacy, confidentiality and ownership, as well as for secure data exchanges with trustworthy parties are of paramount importance. To this end, an increasingly popular alternative to centralised data processing that addresses these concerns is to perform ***in-device data processing*** and employ ***privacy-aware communication schemes*** between devices that do not expose internal user data. Additional perks of this approach include robustness against downtime of centralised infrastructure (e.g., the 2021 Facebook outage had serious ramifications around the globe[190]) and the ability to deploy software and its accompanying AI to places with limited or restricted internet access (e.g., areas

[187] K. Hartnett, To Build Truly Intelligent Machines, Teach Them Cause and Effect (2018): https://www.quantamagazine.org/to-build-truly-intelligent-machines-teach-them-cause-and-effect-20180515/
[188] Y. Gil and B. Selman. A 20-Year Community Roadmap for Artificial Intelligence Research in the US. Computing Community Consortium (CCC) and Association for the Advancement of Artificial Intelligence (AAAI). Released August 6, 2019. arXiv:1908.02624 https://cra.org/ccc/resources/workshop-reports/
[189] A. Kranz, There are over 3 billion active Android device (2021): https://www.theverge.com/2021/5/18/22440813/android-devices-active-number-smartphones-google-2021
[190] A. Asher-Schapiro and F. Teixeira, Facebook down: What the outage meant for the developing world (2021): https://news.trust.org/item/20211005204816-qzjft/

stricken by natural calamities, warzones, regimes where internet activity is monitored). Since devices lie at the "edge" of communication networks (e.g. of the Internet, but the same principles hold true in local networks without external connectivity) the paradigm of (partial or full) in-device data processing is referred to as ***edge computing*** (Figure 53).

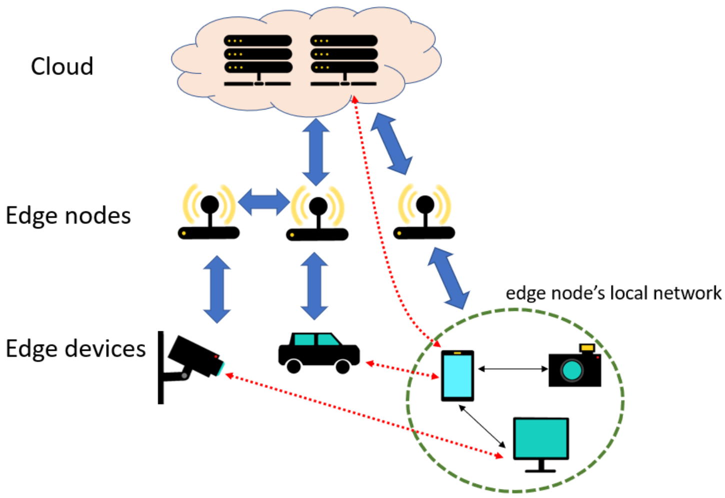


Figure 53: AI at the Edge basic concept. Cloud servers and edge devices can perform local computations or help each other learn through implicit communication paths (dashed red arrows) even if they are not on the same local networks.

Existing research to support AI on the edge can be roughly categorised into three directions, which differ in their degree of autonomy:

On-device inference*.* This aims to deploy AI models with pre-trained parameters to edge devices by replicating inference computations on data residing there. For example, this may take the form of image processing software that performs object recognition or automated tagging without external dependencies. Effectively, inference endpoints are made obsolete by bringing respective computations inside devices, where the latter make inferences autonomously but rely on central services to deploy the trained models. Related research aims to support the deployment of AI on device hardware with new compatibility frameworks (e.g., TensorFlow Lite for mobile GPUs[191]) and create models that fit device resources, for example by supporting low-end hardware or "compressing" the number of trained parameters to reduce memory and processing requirements.[192]

Distributed learning*.* This organises AI model training across several devices by making it independent of where computations are being performed. For instance, a popular paradigm of

[191] TensorFlow Lite: https://www.tensorflow.org/lite
[192] Ogden, Samuel S., and Tian Guo. "{MODI}: Mobile deep inference made efficient by edge computing." {USENIX} Workshop on Hot Topics in Edge Computing (HotEdge 18). 2018.

doing so is *federated learning*[193], which designates one device (e.g., a centralised service) as the trained AI model's host that orchestrates a learning process and lets other devices perform training operations (e.g., gradient calculations) in their own local data. Devices then share back parameter updates to be combined by the orchestrator and are sent copies of the updated model. Distributed learning can be considered as semi-autonomous, because it requires an initialisation process to organise the communication network, but devices run independently. At the same time, it is popular for leveraging the high computational power of relay or cloud servers to learn at scales and speeds unimaginable by centralised computing. An important consideration many distributed systems already address is that devices performing computations may have gathered confidential data, such as medical records, in which case privacy-preserving protocols are employed to make it practically impossible to replicate source data at other devices or the model's host.

Decentralised learning. In this paradigm, fragments of AI models are trained on devices to approximate the outcome of centralised training. Devices do not follow predetermined communication topologies but create unstructured communication links, i.e. which devices communicate is not known at the algorithm design time but only once AI tools are deployed. Existing decentralised learning protocols either consider fixed high-throughput communication overlays (unknown at design time)[194,195] or require the ability to communicate between devices and randomly selected others[196,197], a design referred to as *gossip learning*. In both cases, communicating devices perform model fragment training based on local data and repeatedly average trained parameters between neighbors. Thanks to the conceptual simplicity of this practice, decentralised learning algorithms are often deployed in peer-to-peer communication networks to train model fragments that tightly approximate centralised learning.

Research challenges

AI at the edge provides a promising alternative to existing technological solutions that cope well with the increase in data scale and privacy concerns. However, there remain a lot of open questions over how to support it in real-world scenarios. Below, we outline promising directions that future research can address to support widespread adoption of AI at the edge beyond specialised environments (e.g., distributed learning of cloud or relay servers) to edge devices that see everyday use.

---

[193] McMahan, Brendan, et al. "Communication-efficient learning of deep networks from decentralized data." Artificial intelligence and statistics. PMLR, 2017.
[194] Koloskova, Anastasia, Sebastian Stich, and Martin Jaggi. "Decentralized stochastic optimization and gossip algorithms with compressed communication." International Conference on Machine Learning. PMLR, 2019.
[195] Niwa, Kenta, et al. "Edge-consensus learning: Deep learning on P2P networks with nonhomogeneous data." Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2020.
[196] Hegedűs, István, Gábor Danner, and Márk Jelasity. "Decentralized learning works: An empirical comparison of gossip learning and federated learning." Journal of Parallel and Distributed Computing 148 (2021): 109-124.
[197] Hu, Chenghao, Jingyan Jiang, and Zhi Wang. "Decentralized federated learning: A segmented gossip approach." arXiv preprint arXiv:1908.07782 (2019).

Accountability. Distributed and decentralised AI are trained across multiple devices. Thus, determining accountability is a pressing issue, as there is no single entity responsible for the outcome. Without accountability, even highly accurate AI is difficult to port to high-stakes settings, such as for example in automated medical diagnosis systems[198]. This task is made doubly challenging compared to centralised AI accountability, because different devices could make different conclusions for the same data. At the very least, it is important to dissuade "lazy" practices that lead to harmful (e.g. discriminatory) AI behaviour due to replicating and even accentuating real-world biases (see data heterogeneity challenges).

***Bandwidth limits*.** Distributed and decentralised learning require continuous model gradient or parameter exchanges. At the same time, pre-trained model sizes grow proportionally to the number of their parameters. Hence, more complex machine learning models with billions of parameters may be impractical to deploy through traditional platforms or learning through non-centralised computing. This issue could be projected to the future too, if new computing technologies evolve before communication ones, as has been the trend so far. Preliminary works in this direction address bandwidth limits for gossip learning by performing information exchanges in smaller chunks at the cost of slower learning. Compressing model information during decentralised learning remains an open challenge that can help support more complex models.

***Data heterogeneity*.** Most distributed and decentralised learning approaches consider homogeneous distributions of data across edge devices (e.g., spreading data samples to devices without biases of which device gets which data). However, in practice, devices could differ in terms of the data they collect, for example due to placement of sensors on different physical locations or different preferences of mobile device users. Thus, research must take care to prevent imbalances in the types of local data from becoming biases of local AI model fragments. To make matters worse, these could also be difficult to detect with macro-evaluation (e.g., averaging) of model fragment results.. Gossip learning systemically addresses this challenge by making sure that random pairs of devices exchange parameters, but this comes at the steep cost of requiring constant device availability (accentuating the impact of dynamic behaviour challenges).

***Domain transfer*.** Transfer learning[199] is a widespread paradigm in which trained AI models are repurposed towards different predictive tasks by keeping large chunks of their parameters (e.g. most neural layers) constant and training only the rest. This often helps learn new high-quality models from limited data based on training on similar but larger datasets. For example, a popular practice is to transfer image feature extraction layers of state-of-the-art models to new tasks. This paradigm can be of particular interest for the deployment of pre-trained AI at the edge that can be used to adapt to problems encountered by the device's user. For example, transfer learning can be used to locally turn object recognition software into a recommender

---

198 Yetisen, Ali K., et al. "A smartphone algorithm with inter-phone repeatability for the analysis of colorimetric tests." Sensors and actuators B: chemical 196 (2014): 156-160.
199 Pan, Sinno Jialin, and Qiang Yang. "A survey on transfer learning." IEEE Transactions on knowledge and data engineering 22.10 (2009): 1345-1359.

system that learns from a mobile phone user's stored image to locally refine image web search results, for instance by re-ordering them, without exposing their data to others.

Dynamic behaviour. Current research on non-centralised AI either assumes fixed communication topologies of beneficial characteristics or the ability to randomly communicate with other nodes. However, communication links in the real world may be formed based on the belowmentioned concept of homophily, in which case topologies are fixed but are unlikely to exhibit the desired theoretical characteristics that lead to tight approximation of equivalent centralised model training. At the same time, communication links between devices can be unstable, for example due to users irregularly going online or offline or evolving social relations[200]. Thus, future research needs to address the evolving nature of communication networks (e.g., of peer-to-peer networks), especially those whose links change with rates comparable to learning convergence speed, and in which edge device availability is uncertain.

Homophily. Homophily refers to the tendency of complex network nodes (e.g., social media users) to link with each other based on common attributes[201]. For example, social media friends often have similar hobbies. Thus, device communications (e.g., in peer-to-peer networks) can suffer from a relational type of bias. However, contrary to other risks, homophily, or other relational properties for that matter, can also be leveraged by graph-based AI tools (e.g., graph signal processing, graph neural networks) to make them more accurate. Overall, future research on AI on the edge needs to acknowledge potential homophilous communications and either use this property to improve predictions, or safeguard against potential biases. Notably, leveraging homophily can even support hybrid approaches, where it used by decentralised devices to improve pre-trained inference models

Hyper-parameter selection. This is a concern for decentralised distributed learning and gossip learning approaches only, where there does not exist one central overseer to dictate model hyper-parameters (e.g., the number of neural layers, latent feature dimensions), for example, by comparing alternatives on a validation subset of data. In the case of distributed or gossip learning, hyper-parameters can be selected a-priori through experiments on similar datasets, but may not port well to new data once deployed in the wild. An elegant alternative would be for decentralised AI to also learn its hyper-parameters on-the-fly through additional decentralised processes. When doing this, it is important to create hyper-parameter selection protocols of low computational complexity that do not require untenable training times.

Unequal device resources. The computing capabilities of devices on which AI is deployed may vary and even be unknown at design time. Thus, a promising trend is to create adaptive models that can make the best use of device resources, for example by providing many models for in-device inference. In case of gossip learning, making use of many devices to train fragments of models means that training is lightweight enough to be supported by even older devices.

[200] Berta, Árpád, Vilmos Bilicki, and Márk Jelasity. "Defining and understanding smartphone churn over the internet: a measurement study." 14-th IEEE International Conference on Peer-to-Peer Computing. IEEE, 2014.
[201] McPherson, Miller, Lynn Smith-Lovin, and James M. Cook. "Birds of a feather: Homophily in social networks." Annual review of sociology 27.1 (2001): 415-444.

However, resource allocation may still be unequal in terms of available bandwidth.[202] Overall, research on resource usage needs to make sure that AI on the edge is not as weak as the lowest computing capabilities of devices expected to run computations.

Societal and media industry drivers

Vignette: Debunking fake news under an authoritarian regime using AI at the edge

Ann, Bob and Cale are journalists stationed inside the territory of an authoritarian regime. The regime closely monitors internet activity and keeps producing disinformation content in order to spread propaganda in its populace. All three journalists come across fake media on a daily basis, although these are only a small portion of the regime's continuous efforts at disinformation, i.e., there are more pieces of fake media they do not come across. The journalists aspire to support AI tools that, based on their annotations, would learn to identify more pieces of disinformation to warn the populace. However, due to ongoing monitoring of internet activity, doing so runs the risk of them being caught.

Luckily, all three journalists are users of a decentralised fake media detection platform. In this, the users annotate (i.e. tag) media examples as fake or not and these annotations are fed to decentralised learning algorithms to learn to distinguish fake media similar to the annotations. The platform runs on a peer-to-peer network (e.g. that already circumvents part of monitoring by encrypting its communication) that maintains privacy by fully obfuscating how users contribute to the fake media detection algorithm. Furthermore, the platform is designed to send user data only to trusted others.

Thus, the journalists can feel safe in providing high-quality annotations, which decentralised algorithms will then collectively process so that the devices of all platform users would hold fragments of predictive models that learn to distinguish whether viewed media content is fake or not.

Future trends for the media sector

Edge computing can be a game changer in the way the media sector deploys AI models to enrich media content with metadata and develop new user experiences. We highlight some of these opportunities with an eye to AI4Media use cases:

- **Create collectively trained AI models** that process and learn from continuously generated real-world data. For instance, these models could interact with the users to obtain feedback on classification or recommendation goals; if only a portion of users are willing to manually annotate data meant for their own consumption, then all devices can make use of this information. We expect the increased privacy of AI to encourage users to engage in this way, perhaps through software design opportunities (e.g., like and dislike buttons in mobile applications) that would not be possible for fully centralised systems.

[202] Musaddiq, Arslan, et al. "Reinforcement learning-enabled cross-layer optimization for low-power and lossy networks under heterogeneous traffic patterns." Sensors 20.15 (2020): 4158.

- **Reduce development and upkeep costs of AI tools** by making use of the combined computing power of their users' devices to run calculations, i.e., media companies can avoid inference costs and -in the cases of distributed or decentralised learning- model maintenance costs (e.g., training with new data and re-deploying). This also means that AI training can become more environmentally-friendly, as the already running resources of edge devices are used.
- **Ensure data privacy and ownership** by not allowing data to leave user devices. This can help significantly promote trust of content users.
- **Create highly personalised media applications** that take into account many aspects of user lives that would be difficult to gather and get back with centralised architectures.
- **Personalise AI** based on local user feedback.
- **Help fight disinformation** by creating collectively managed environments that make use of AI without the tampering of overseers.
- **Perform ongoing training** that quickly adapts to changes in real-world data in safety-critical systems. This is particularly useful against adversarial attacks that aim to fool AI tools by circumventing current learned models. For example, decentralised learning could provide protection against evolving disinformation and deepfake techniques by leveraging new content flagged by users. Users could flag new types of fake content and decentralised AI could immediately integrate this information in model training to also flag similar content. This would be achieved without waiting for software updates (bearing new versions of models) that take too long to deploy and could help stop disinformation attacks long before they reach a critical mass of users to become popular.
- **Allow smaller media organisations to compete** on the AI front without requiring expensive machinery or extensive data collection processes.

Goals for next 10 or 20 years

In the next 10 or 20 years, AI at edge will be able to make use of most data generated by edge devices to capture multifaceted aspects of people's lives without violating their privacy. Thus, highly personalised and well-scaling AI will be able to enrich a new generation of human-machine interactions, where trainable models (including those used in media applications) learn from the collective experience of their target audiences and support learning tasks that are not feasible by existing centralised computing.

5.7 Bioinspired learning

Contributors: *Nicu Sebe (UNITN), Wei Wang (UNITN), Cigdem Beyan (UNITN), Marco Formentini (UNITN), Kasim Sinan Yildirim (UNITN), Enver Sangineto (UNITN)*

Current status

During the last decade, the multimedia and computer vision research communities have witnessed the revolution brought by deep artificial networks, which are inspired by the biological visual system. However, there still exist discrepancies between deep networks and biological ones. To advance beyond the current deep learning scheme, one needs to re-think and re-model current deep network architectures by reverse engineering of the human visual system including the cognitive patterns, neuron connections, and the capability of continual learning.

Deep neural networks take inspiration from the human brain. The quick development of computing platforms (*e.g.*, Graphics Processing Unit[203] (GPU) and Tensor Processing Unit[204] (TPU)) provides strong computing power and paves the way to further development of AI. It has been shown for object recognition[205,206,207], tracking[208], image labelling[209] and other fields that features learned for a specific problem using deep convolutional neural networks (CNNs) show much better performance than traditional machine learning approaches. Instead of splitting the feature and classifier learning processes, CNNs supports *end-to-end* learning of the feature extractor and classifier simultaneously. The *end-to-end* learning mechanism enables CNNs to learn task-specific features automatically. The very first CNN model is LeNet[210] proposed in 1998. Eventually, after nearly 15 years, with the help of powerful computing platforms (GPU, TPU), ground-breaking models winning the ImageNet Large Scale Visual Recognition Challenge[211] were established, including AlexNet[205] in 2012, VGG19[209] & GoogleNet[206] in 2014 and ResNet[207] in 2015. Since then, no significant progress has been made and the new models are usually an ensemble of previous models.

---

[203] J. Sanders and E. Kandrot. CUDA by example: an introduction to general-purpose GPU programming. Addison-Wesley Professional, 2010.
[204] N. P. Jouppi, et al. In-datacenter performance analysis of a tensor processing unit. In International symposium on computer architecture, pages 1–12, 2017.
[205] A. Krizhevsky, I. Sutskever, and G. E Hinton. Imagenet classification with deep convolutional neural networks. NeurIPS, 25:1097–1105, 2012.
[206] C.n Szegedy, et al. Going deeper with convolutions. In CVPR, pages 1–9, 2015.
[207] Kaiming He, et al. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.
[208] N. Wang and Dit-Yan Yeung. Learning a deep compact image representation for visual tracking. In NeurIPS, 2013
[209] K. Simonyan and A.Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[210] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.
[211] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, pages 248–255, 2009.

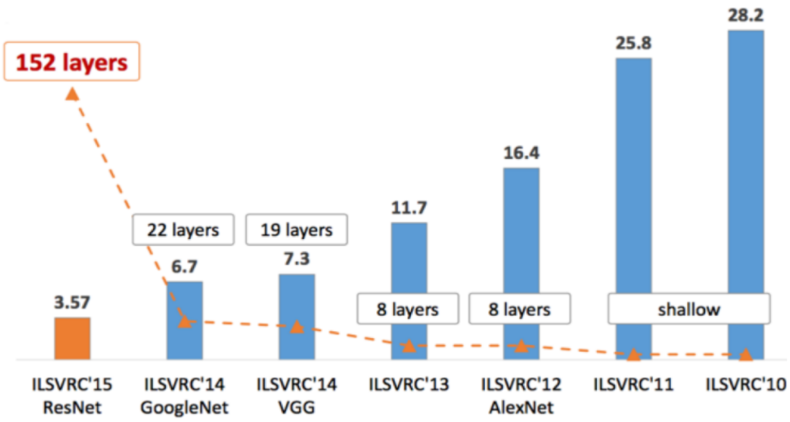


Figure 54: Evolution of CNN architectures [212]. The object recognition rate (%) of each network is depicted as a bar.

The **ImageNet competition** made a large contribution to the development of AI. ImageNet is a large visual database designed for visual object recognition research with more than 14 million images. ImageNet runs an annual contest, *i.e.*, the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), where software programs compete with each other to correctly classify and detect objects and scenes. Both the industry and research communities devote their energies in the ILSVRC competition by making their CNN models deeper and wider. With the popularity of CNNs, giant companies such as Google, Facebook, Microsoft, and Amazon also take active part in the development of CNNs. Figure 54 above shows the evolution of CNN architectures. The bar represents the object recognition error (smaller means better). We can observe that as the network goes deeper, the performance keeps improving. With the network depth gradually increasing from 8 to 152, the recognition error drops from 16.4% to 3.6%. Note that depth increase does not necessarily lead to the increase of the number of parameters. However, in the learning stage, the CNN model parameters and their intermediate features and gradients need to be stored. Therefore, more memory is required to train the very deep CNN models, and for example a 152 layer deep ResNet could drain out the memory and computing power of one individual GPU. Later on, more complicated CNN models such as DenseNet[213] needed to run on GPU clusters in order to have enough memory to host the model. In other words, the limited memory and computing power of the GPU have restricted the CNN models to go even deeper. Therefore, how to optimise CNN architecture to achieve better performance without increasing the memory and computing power consumption will be the main focus of the AI community in the future. The solution lies in the ***reverse engineering of the human vision system***.

To further advance the development of AI, computer science research scientists have tried to mitigate the gap between artificial and biological neural networks. For instance, *Geoffrey Hinton*,

[212] Image source: OpenGenus, Evolution of CNN Architectures: LeNet, AlexNet, ZFNet, GoogleNet, VGG and ResNet https://iq.opengenus.org/evolution-of-cnn-architectures/
[213] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In CVPR, pages 4700–4708, 2017.

the pioneer of AI, has published two open access research papers[214,215] on the theme of capsule neural networks which can be used to better model hierarchical relationships. This approach is an attempt to mimic the organisation of biological neural structures more closely. In the meanwhile, from the neuroscience research community, many works on biologically inspired artificial neural networks have been developed, such as the spiking neural networks[216]. All these works share a similar vision, *i.e.,* **bringing more neural realism into deep networks and reducing the differences between the artificial and biological neural networks**. The human vision system perceives the outside world in a more efficient way which is quite different from CNN models. The main differences lie in **i) recognition patterns, ii) network topological structures, and iii) the memory mechanism.**

Research challenges

To reduce the gap between deep neural networks and biological ones, we need to take a closer look at their differences from three aspects: i) Is it possible to build an artificial neural network that has the same cognitive pattern and behaves in the same way as the biological human visual system? ii) Can we design a smarter artificial neural network by exploring diverse neural network topologies that exist in the human brain? iii) Is it possible to enable the artificial network to have a similar continual learning ability as a human?

To answer these three questions, the solution is to i) simulate the dual stream cognitive pattern of human vision; ii) model the diverse topological structures of biological neuronal circuits; and iii) explore the possibility of continual learning to reach a quite similar deep neural continual learning ability as a human.

One fundamental principle in human cognitive patterns is that the human visual cortex possesses two distinct streams i.e., ventral and dorsal as shown in Figure 55[217]. The ***ventral stream*** ('what' path- way) is involved in high-level perception[218] (e.g., object/scene recognition) while the ***dorsal stream*** ('where' pathway) is involved in spatial cognition. The two streams correspond to the classic definition of computer vision proposed by David Marr[219] which is to look at 'what' is 'where'. In the context of computer vision, 'what' denotes object recognition (object vision) and 'where' refers to 3D reconstruction and object localisation (spatial vision)[220]. This paradigm guides the research in computer vision, but the spatial and object vision tasks are usually studied independently. However, most deep networks, such as ResNet designed for either classification, segmentation or object detection have focused on designing one-shot

[214] S. Sabour, N. Frosst, and G. Hinton. Dynamic routing between capsules. arXiv preprint arXiv:1710.09829, 2017
[215] G. Hinton, S. Sabour, and N. Frosst. Matrix capsules with EM routing. In ICLR, 2018.
[216] W. Gerstner and W. Kistler. Spiking neuron models: Single neurons, populations, plasticity. Cambridge university press, 2002.
[217] D. Milner and M. Goodale. The visual brain in action, volume 27. OUP Oxford, 2006
[218] L. Cloutman. Interaction between dorsal and ventral processing streams: where, when and how? Brain and language, 127(2):251–263, 2013
[219] D. Marr. Vision: A computational investigation into the human representation and processing of visual information. 1982.
[220] M. Mishkin, L. Ungerleider, and K. Macko. Object vision and spatial vision: two cortical pathways. Trends in neurosciences, 6:414–417, 1983.

methods, that is, algorithms that take an image as input, process it, and return an output without any feedback loop. This is in contrast with what we know about the human vision system where the two streams work collaboratively for the perception of the outside scene.

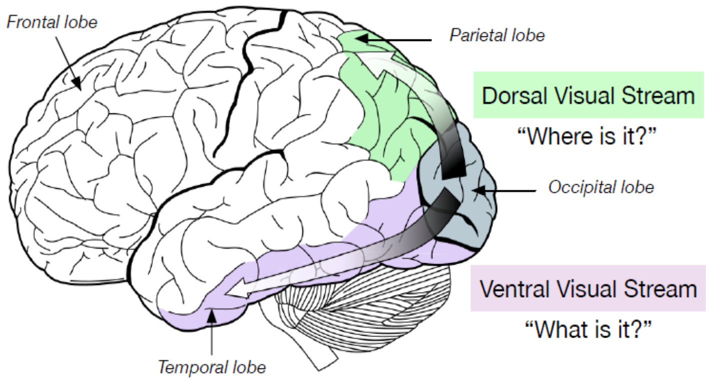


Figure 55: Human vision system. The dorsal stream determines 'where is it' while the ventral stream determines 'what is it'. The two streams originate from a common source in the visual cortex.[221]

This dichotomy shows that it is very necessary to design a **dual stream recognition pattern for recursive coarse-to-fine perception**: the ventral network is in charge of high-level perception while the dorsal network retrieves the memory associated with the ventral network and attends to salient objects to refine the perception repeatedly. In this way, the dual stream networks compose a cyclic loop and refine the perception progressively.

Second, the deep network neurons are activated non-linearly in the style of human neurons, but some important topological structures are ignored, such as the recursive connection. Besides, the biological neural network has very complex local topological structures as shown in Figure 56. Diverse combinations of these structures lead to various networks with diverse global structures. Many classical modules in artificial networks can be viewed as simplified duplicates of the biological neuronal circuits. For instance, the residual connection in ResNet can be viewed as a special case of a parallel neuronal circuit[222] (as shown in Figure 56(4)), in which there is only one branch running in parallel with the main stem. Besides, the inception module in GoogleNet can be viewed as the combination of diverging and converging circuits as shown in Figure 56 (1) & (2).

This limited amount of artificial modules with simplified local topological structures has helped the deep models to achieve remarkable performance in computer vision tasks. We can push it one step further through **modelling the neural circuits with rich topological structures**, and thus to design more powerful artificial modules which can model more complex functions. To bring this into reality, we need to rely on the strong computing power of servers nested with GPU clusters or TPUs. In the meanwhile, AI has defined its own advantages *w.r.t.* its electronic computing platform, such as the speed, reconfigurability, parallelisation, and scalability. It has

[221] Image source: Wikipedia - https://en.wikipedia.org/wiki/File:Ventral-dorsal_streams.svg
[222] K. Saladin and R. McFarland. Human anatomy, volume 3. McGraw-Hill New York, 2008.

the potential to surpass the cognitive intelligence it tries to mimic. For instance, our eyes can only orient our gaze to one salient object at a time, while the computer can process different regions in parallel.

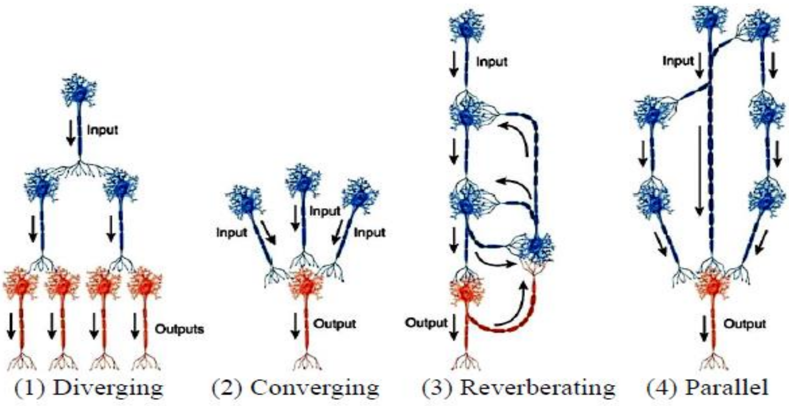


Figure 56: Biological neuron circuits.[223]

Moreover, in the real world, we are exposed to continuous streams of information. To adapt to the changing environment, we are able to learn multiple tasks from dynamic data distributions in a continuous manner. The ability to continually learn over time by accommodating new knowledge while retaining the previously learned one is referred to as ***continual or lifelong learning***. In the context of AI, it means being able to smoothly update the artificial network to perform more tasks but still being able to *re-use* and *retain* knowledge that has been previously learned without forgetting it.

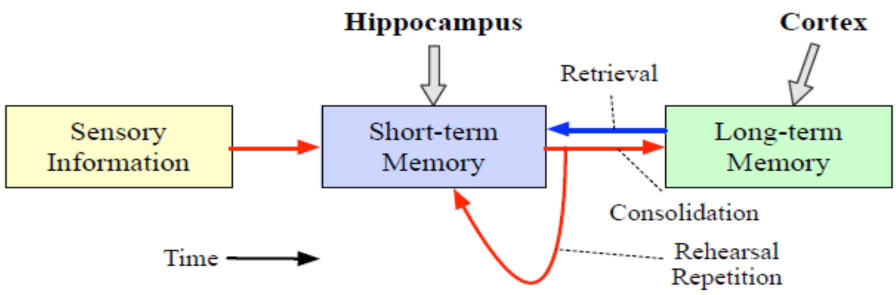


Figure 57: Dual memory system.

By studying *Henry Molaison*[224] who was unable to form new memories after removing his hippocampus to treat epilepsy, neuroscientists revealed the dual memory mechanism in our brain which may play a vital role in continual learning. It is believed that the hippocampus has a short-term memory and it is involved in rapid learning of new tasks. It encodes sparse representations to minimise interference. In contrast, the neocortex has a slow learning rate

[223] Image taken from: Curtis DeFriez , "Chapter 12 Nervous Tissue" at https://slideplayer.com/slide/5964323/ (slide 89)
[224] The Brain Observatory, Deconstructing Henry, https://www.thebrainobservatory.org/projecthm

and is involved in learning generalities by building overlapping representations of the learned knowledge. As shown in Figure 57, knowledge is transferred from short-term memory to the long-term storage memory via knowledge consolidation[225]. The long-term memory can then be recalled and reconsolidated[226] due to *neural plasticity*[227], meaning that it can be adapted by acquiring, refining, and transferring knowledge across multiple domains[228]. The cooperation between hippocampus and neocortex is key to learn high level regularised concepts and memory, but the exact mechanisms are still not yet completely understood. To address the catastrophic forgetting problem in continual learning, one solution is to **model the dual memory mechanism**. Specifically, the regularised concepts that are shared across tasks will be represented by an attribute dictionary, which will be stored in the long-term memory module. The short-term memory module is allowed for gradually forgetting and it can keep learning new attributes and transfer them to the attribute dictionary. By referring to the dictionary, novel objects can be easily represented. For instance, with the following attributes: 'shape like horse' and 'black-white' 'stripes, we are able to represent 'zebra'.

Societal and media industry drivers

Vignette 1: AI-enabled web-camera for highly realistic virtual interactions in the Metaverse

Marco is having a meeting with his friends. Because of the pandemic, they could not meet onsite. Therefore, they have to stay in their own room and join an online 3D virtual meeting room in which they project themselves as 3D-realistic avatars in the virtual 3D spaces. The webcam has the same cognitive pattern as a human and it can synchronise the avatar and Marco when he talks, moves, and interacts. When humans observe a scene, their eyes look into different directions and orient their gaze to the location where a visual object has appeared. For example, when observing a still face image, our eyes undergo a saccadic movement and re-target to salient regions (see Figure 58). With an AI-enabled web-camera, the moving trajectory of Marco's pupils can be tracked in real time to know his gaze orientation. Together with the sensors that detect his head movement, Marco's views in the virtual environment can be changed automatically. With the help of such a cognitive pattern, the webcam can capture all of Marco's movements and make the avatar come to life.

[225] Y. Dudai. The neurobiology of consolidations, or, how stable is the engram? Annu. Rev. Psychol., 55:51–86, 2004.
[226] K. Nader, G. Schafe, and J. Le Doux. Fear memories require protein synthesis in the amygdala for reconsolidation after retrieval. Nature, 406(6797):722–726, 2000.
[227] E. Fuchs and G. Flugge. Adult neuroplasticity: more than 40 years of research. Neural plasticity, 2014, 2014.
[228] A. Bremner, D. Lewkowicz, and C. Spence. Multisensory development. Oxford University Press, 2012.

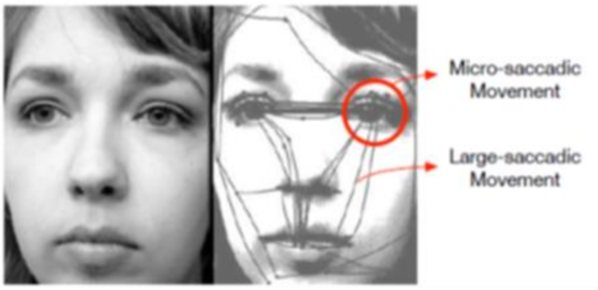


Figure 58: Saccadic eye movements.[229]

Vignette 2: Playing video games with smart AI companions

John is wearing a virtual reality camera and is playing a game with an AI agent. After being familiar with the rules, John can always win the game with the same strategy and he feels bored. He then selects a smarter AI agent with the ability of continual learning. The smarter agent can keep learning from mistakes without forgetting the experience learned previously. Similar to a human, the AI agent can adapt itself by acquiring, refining, and transferring knowledge. After each game, the AI agent will summarise the good strategies of good actions and bad strategies of bad actions. These strategies and experiences are all stored in the long term memory unit of the agent and it can behave more like a human. Therefore, John must come out with new strategies to win the game. Moreover, the AI agent can keep evolving automatically and become smarter by playing games with another smart AI agent. As such, John will not be bored anymore.

Future trends for the media sector

The progress of neuroscience and deep learning theory[230], together with the development of powerful computing platforms (from CPU to GPU and TPU), make up the basis for the development of AI. The next evolution of social networking is to help bring the Metaverse to life. One of its main characteristics is the use of 3D spaces that can let one socialise, learn, collaborate and play in new ways. To bring this to life, 3D modelling is the key. Therefore, it is very important for AI to understand the 3D world so that the 3D objects could be well reconstructed in the virtual 3D spaces.

As humans, we naturally have the ability to extract 3D information using our dual stream visual cognitive pattern. Figure 59 shows the dual stream visual cognitive pattern of a human. The ventral stream recognises that the image shows a pair of shoes (left). Next, the dorsal stream network will pay attention to the shoes and extract the 3D location (middle) and shape information which is represented by the segmentation mask (right). By mimicking human visual cognitive patterns, AI can better understand and reconstruct the 3D scenes and improve the user experience.

[229] Image source: Wikipedia - https://en.wikipedia.org/wiki/Saccade#/media/File:Szakkad.jpg
[230] Adam H Marblestone, Greg Wayne, and Konrad P Kording. Toward an integration of deep learning and neuroscience. Frontiers in computational neuroscience, 10:94, 2016.

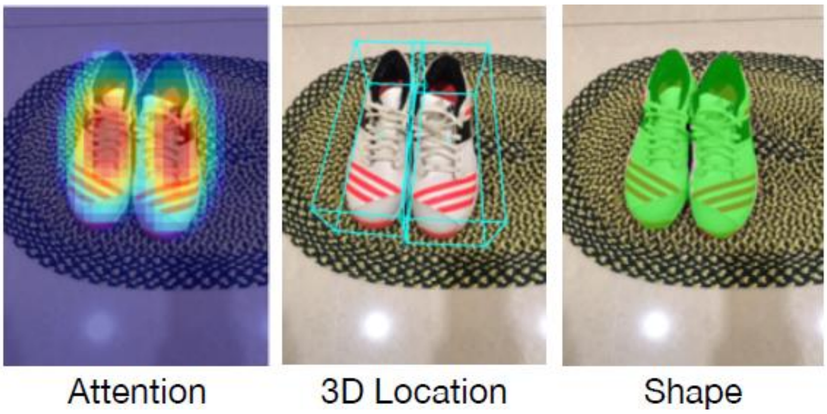


Figure 59: Dual stream recognition example.[231]

Besides, social media has a huge amount of content in the format of videos, images, texts, etc. It is very necessary to summarise these contents so that we can better categorise them and present them to users. Even though we already have many deep neural networks to do the job, they are usually only good at one media format. For instance, the network may be good at action recognition in videos, but not good at face recognition. Therefore, it is very necessary to design more powerful deep neural networks. One potential solution is to model the various topology structures of the neurons in the human brain. In this way, the deep neural networks may be good at multiple tasks across different media formats. Moreover, the world is changing dynamically, and new knowledge keeps coming out. It is also very important for the deep neural networks to learn continuously to adapt themselves by acquiring new knowledge, refining them, and transferring them across multiple domains.

Goals for next 10 or 20 years

Instead of focusing on pursuing high performance and fast speed relying on the target computing resources (*e.g.*, GPUs, embedded devices), in the future, AI will copy the processes that underlie the way a biological system thinks and remembers and will take them one step closer to a real living biological system. By rethinking the artificial networks from the view of the biological system and the reverse engineering of the human visual system based on the basic theories discovered by neuroscientists, AI will behave in a more similar way as humans.

The long-term vision of AI is to enable artificial networks to process and perceive visual information like the human visual system and to learn knowledge continually like the human brain (next 5-10 years). By pushing the artificial network closer to the biological one, it might be easier to integrate artificial networks into biological ones. This will also have great potential to replace damaged brain sectors or make our brain more powerful by planting artificial circuits into it (next 10-20 years).

[231] Image source: MediaPipe - https://google.github.io/mediapipe/solutions/objectron.html

5.8 Quantum computing

***Contributors**: Artur Garcia (BSC)*

Current status

With Moore's Law failing and anticipated to flatten by 2025[232], three different evolution paths have been predicted for post-exascale systems. One path considers the development of ever more specialised architectures and advanced packaging technologies that arrange existing building blocks in new ways (next 10 years after exascale); another path considers the development of CMOS-based devices that extend into the third, or vertical, dimension and on improving materials and transistors that will enhance performance by creating more efficient underlying logic devices (next 20 years); the third axis represents opportunities to develop new models of computation, such as neuro-inspired or quantum computing, which solve problems that are not well addressed by digital computing. Exploiting the microscopic properties of matter, as described by quantum mechanics, opens the door for a completely novel formulation of information processing tasks. ***Quantum computation***[233] emerges as a revolutionary technology, with the expectation to have deep impact on fundamental and applied aspects of computation, ranging from exponential speedups of complex computations, to a reduced cost and power consumption compared to major supercomputers.

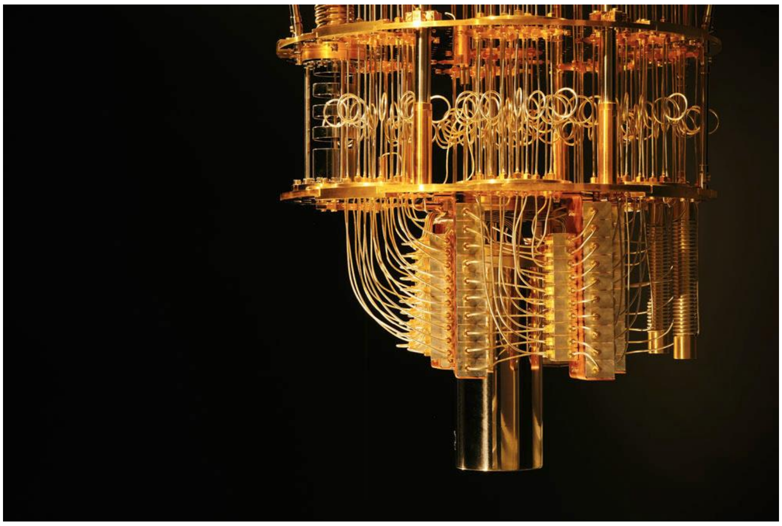


Figure 60: Quantum computer [234].

[232] J. Shalf. "The future of computing beyond Moore's law." Philosophical Transactions of the Royal Society A 378, no. 2166 (2020): 20190061.
[233] M. Nielsen and I. Chuang, Quantum Computation and Quantum Information, Cambridge University Press (2000)
[234] Image source: IBM Newsroom - https://newsroom.ibm.com/media-center?keywords=quantum#gallery_gallery_0:21747

Since quantum computers are neither cheap nor easy to build, classical simulation is a valuable method for efficient simulation of quantum algorithms. For example, quantum devices are affected by noise which currently is a significant limitation to enhance their capabilities. Classical simulation tools are a must to understand noise sources and improve the performance of quantum algorithms. These tools overcome the fundamental limitation of measurement of quantum states, offering a deep insight into how a quantum computer works. Furthermore, classical quantum circuit simulation gives in-depth information about how future quantum machines will behave, reducing the cost to build and maintain them. Thus, improving classical simulation tools will help to better understand and to optimise quantum devices. Also, these simulations play a crucial role in the development of novel quantum algorithms.

However, these advantages come at a significant cost. The simulation of a quantum circuit is exponentially expensive with its size, thus requiring the power of **High-Performance Computing** (HPC) technologies even for small instances[235]. For example, one work[236] reports the usage of 196 TBytes (the whole available nodes' memory) for a 42 qubits circuit simulation while another[237] reports HPC simulations of 49-qubit hard random quantum circuits (RQC) on 4,600 out of 4,608 nodes of the Summit supercomputer, the second largest supercomputer in the world according to the Top500 list of June 2021[238]. Hard instances emerge as a combination of properties of the quantum circuit, more importantly its width (N, number of qubits) and depth (D), its simulation complexity depends on its product NxD and also on the resulting connections between qubits. While easy instances can be solved with current approaches, hard instances do not fit in current HPC systems making the simulation unfeasible. The use of HPC technologies is a crucial element in the design of new quantum devices, which will take a huge importance in the next few years to improve their quality.

Research challenges

Recent efforts to establish the computational power of quantum computers relate their capabilities to classical computers using computational complexity, where one classifies a given problem based on the efficiency with which we are able to solve it. There exist quantum algorithms for classical problems that are much better than any known classical algorithm[239,240], even if they are too complex for the current quantum devices. Recently, proposed Noisy Intermediate-Scale Quantum (NISQ) devices have approached the breaking point of quantum

[235] Y. Zhou, M. Stoudenmire, and X. Waintal. "What limits the simulation of quantum computers?." Physical Review X 10, no. 4 (2020): 041038
[236] G.G. Guerreschi, J. Hogaboam, F. Baruffa, and N. Sawaya. "Intel Quantum Simulator: A cloud-ready high-performance simulator of quantum circuits." Quantum Science and Technology 5, no. 3 (2020): 034007
[237] Villalonga et al. "Establishing the quantum supremacy frontier with a 281 pflop/s simulation." Quantum Science and Technology 5, no. 3 (2020): 034003
[238] TOP500 The list (2021): https://www.top500.org/lists/top500/2021/06/
[239] L. Grover. "A fast quantum mechanical algorithm for database search". In:STOC '96.1996.
[240] P. Shor. "Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer". In:SIAM Journal on Computing 26.5 (Oct. 1997), pp. 1484–1509. issn: 1095-7111.doi:10.1137/

advantage[241,242], where they perform simulations that cannot be done efficiently on a classical computer. In general, finding the problems that offer such an advantage on quantum computers and why is at the core of theoretical research in the field. These quantum devices have been implemented using different quantum technologies, and quantum advantage has been shown for superconducting circuits and photonic devices.

Simulation of quantum circuits has been recently leveraged to support the claims of supremacy of quantum computations over their classical counterparts. Afterwards, these claims have been revisited, showing the need of a general effort to establish the full potential of HPC systems to simulate quantum circuits. The use of tensor networks for quantum computing simulation is relatively new, however already identified as the leading approach. Igor L. Markov and Yaoyun Shi first acknowledged the power of tensor networks for quantum computing simulation and the tensor contraction the more common and complex operation[243]. While there are implementations of software simulation of quantum circuits using tensor networks, there is a general lack of methodologies to use HPC systems to simulate very large tensor networks.

A major research challenge is to define in a solid way the computing capabilities of quantum computers against current technologies in practical problems. These problems have to be selected to be of practical interest, yet needing a large computational power. Among these applications, one identifies machine learning related operations as those interesting to boost with advanced technologies such as quantum computing. However, finding a quantum algorithm suitable for these tasks is just a first step in a series of technological challenges. The final one, namely constructing an operational quantum computer in the lab, is a major challenge in our current manipulation of systems in the microscale.

Societal and media industry drivers

Vignette: Analysing gigantic image datasets with the help of quantum computing

Lisa is a researcher in a small AI startup analysing an enormous dataset of images for object recognition to build an AI model for a public service media organisation that wishes to exploit their vast audiovisual archives. Unfortunately, the available computing resources are running out of free space. However, a subroutine of her pipeline has been improved with a quantum version. This variant allows a large speedup assuming data can be loaded efficiently to a quantum device. Fortunately, Lisa has access to a cloud account for a nearby research institute where classical and quantum machines run together in hybrid systems. She uploads the data to the hybrid cloud device and runs the pipeline. The novel quantum algorithm performs very well and the boost is significant, despite the overhead loading and unloading the data to the quantum device. More tests will be required to assess the scope of this speedup in the following datasets,

[241] Han-Sen Zhong et al. “Quantum computational advantage using photons”. Science 370.6523(2020), pp. 1460–1463
[242] F. Arute et al. "Quantum supremacy using a programmable superconducting processor." Nature 574, no. 7779 (2019): 505-510
[243] I.Markov and Y. Shi. “Simulating quantum computation by contracting tensor networks”, SIAM Journal on Computing, 38(3):963-981, 2008

as these keep growing in size. Fortunately, due to low energy requirements and short duration of the quantum computations, the dataserver runs on green energy and is not raising prices since last year.

Future trends for the media sector

A major resource in data manipulation using AI is computational power. Dataset size is a major indicator of the complexity, but also the intrinsic hardness of a particular problem can be a major hurdle to reach a good solution. Reaching to alternative computing technologies is becoming a popular study for fields reaching a saturation of their available resources.

For providing better solutions with quantum devices performing AI tasks, with processing capabilities for large datasets of text, images, and media in general, one needs to develop novel quantum algorithms, i.e. algorithms intended to run on a quantum device. These novel algorithms are hard to formulate, and the community has been only able to produce a few. However, with more implementations of real quantum computers, one expects an increasing interest of researchers for the possibilities of these novel formulations.

Among other interesting avenues of research, quantum formulation of classical algorithms allows an alternative approach to well-known numerical methods. However, this novel point of view has provided important insight, and even improvements, over these methods[244]. With these quantum inspired algorithms, with clear applications to data science, one may benefit from the power of the quantum formulation of an algorithm, without the need to execute them on a real quantum algorithm.

Following results of benchmarking classical and quantum computations for well-known problems, we can establish a direct comparison between these technologies from the perspective of the resources used. These include running time, memory, but also energy consumption and fabrication costs. The expected consumption of novel supercomputers exceeds by orders of magnitude those of quantum computers. This opens a new opportunity of cheap computing power in some specific tasks (e.g. NLP), which may be relevant even in conditions on which classical computers offer an advantage in computing power, but a large disadvantage in any other consumed resources.

Goals for next 10 or 20 years

The major goal of the field of quantum computation, and a necessary condition in order to provide real quantum solutions, is the production of functional quantum computers providing a large number of qubits. This requirement allows the execution of error correction over the quantum system which is plagued with experimental errors. With this quantum computer, one is then able to further develop, test and execute novel quantum algorithms.

Quantum computing development in the next years is expected to be a hot topic and we expect many novel advancements. IBM among others (such as Google, Rigetti or IonQ) is one of the

[244] E. Tang, Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions, Phys. Rev. Lett. 127, 060503 (2021)

main actors developing quantum hardware and their predictions for the next years include devices with 1000+ of Qubits starting at 2023 (Figure 61). Similar projections by other actors suggest the availability of 1M+ of Qubits in the mid-term. With these resources one expects to offer fully capable quantum computers.

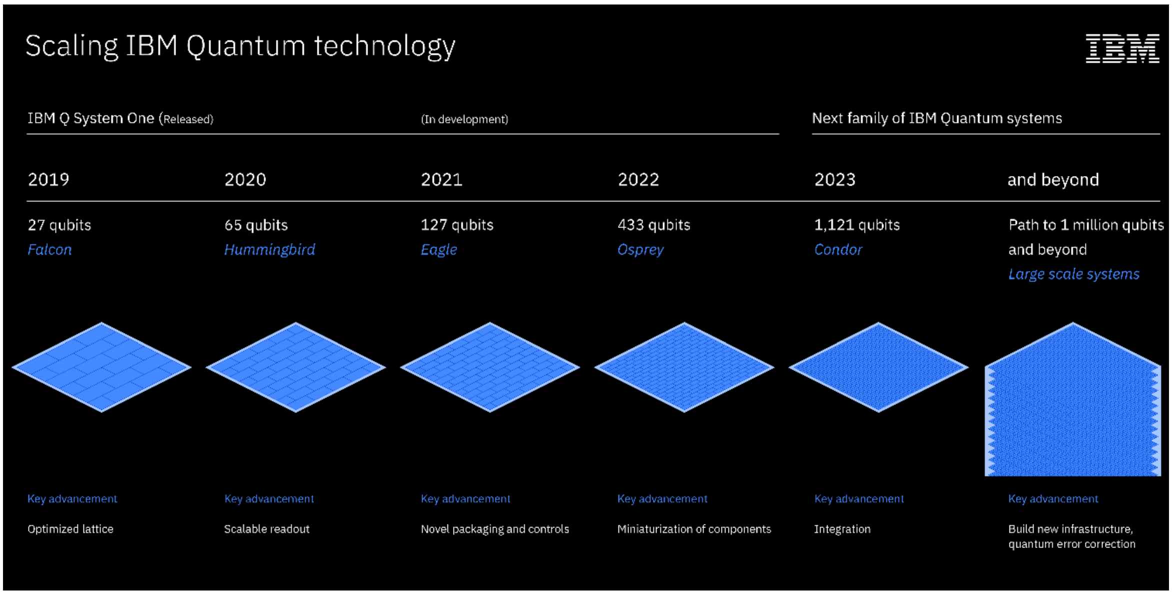


Figure 61: Infographic - IBM's Roadmap for Scaling Quantum Technology [245].

Having a working quantum algorithm is a starting point for major advances in computational applications. However, a careful benchmark of a new computational device has to be performed to assess its utility in different tasks. The field of benchmarks for quantum devices will have to provide answers to which tasks will benefit from the development of real quantum computers.

Finally, while we are still far from having full powered quantum computers in the lab, hybrid architectures can be built from the integration of classical and quantum technologies. The integration of a noisy quantum device with classical computers (even in HPC environments) may take advantage of a limited selection of quantum features under the robust setup of an established classical methodology.

[245] Image source: IBM Newsroom - https://newsroom.ibm.com/media-center?keywords=quantum#gallery_gallery_0:21737

6 AI for multi-media applications: what does the future hold?

In section 3, we analysed a selection of roadmaps, surveys, and articles focusing on media AI applications but also on more general AI technology trends. Based on the results of this analysis, in section 3.3 we identified the most prominent AI-enabled applications and technologies of interest for the extended media industry.

In this section, we focus on the analysis of AI technologies and applications for multimedia analysis, including *multimodal knowledge representation and retrieval, media summarisation, automatic content creation*, *affective analysis*, *NLP applications* and *content moderation*. The aim is to offer a clear overview of the current status of these technologies, the drivers and challenges for their development and adoption, and their future outlook in order to provide some insights on how these technologies could help transform the media and entertainment industry.

A separate subsection is dedicated to each technology, following the same format that was adopted for section 5 (see the introductory paragraph of that section). Each subsection includes the following parts:

- ***Current status***, providing general information about the technology and its current status;
- ***Research challenges*** for the development and adoption of the technology;
- ***Societal and media industry drivers***, focusing on what potentially drives the adoption of this technology in the media and entertainment industry by providing short ***vignettes***, i.e. examples that effectively showcase how AI innovations could impact society and business;
- ***Future trends for the media sector***, presenting the potential applications of these technologies in different sectors of the media industry; and
- ***Goals for next 10 or 20 years***, summarising insights and predictions for the further development of these technologies and their application in the media.

6.1 Cross-modal and multimodal representation, indexing and retrieval

Contributors: *Frederic Precioso (3IA-UCA), Lucile Sassateli (3IA-UCA)*

Current status

Multimedia content covers all modalities (visual, audio, text, etc.) and each modality is carrying its own specific piece of information. To gather all these pieces together and explore the whole information all at once, each modality requires to be represented in a common description space where they can all be combined and compared. This new digital representation has to preserve the essence of the information carried by the different modalities.

For decades, each modality representation was addressed with dedicated techniques designed by experts in acoustic for audio, in image processing for images and video, in linguistics for texts, etc. Once multimedia content was converted in these new representations, one had to define the organisation and the storage, i.e. the indexing, of this content such that information retrieval in all the multimedia content processed would be faster and easier.

During the last decade, the methods for representing multimedia content have been revolutionised by the emergence of deep representation learning. The paradigm of data representation has moved from precisely hand-crafted feature extraction to learning the representations as parts of training deep neural network architectures from data. These advances have particularly impacted how to efficiently represent audio data[246], visual data,[247,248,249] video data,[250] or textual data[251].

These new efficient methods have allowed to drastically reduce the transfer of multimedia analysis models from research labs to the market, simplifying the design of new multimedia retrieval systems even by non-experts. By leveraging most challenges of the last decades on multimedia content representation, these methods caused a shift of focus on more challenging tasks, moving from multimodal analysis to cross-modal analysis. In cross-modal analysis, only one modality is exploited to retrieve the content information in all modalities. Visual Question Answering (VQA), i.e. open-ended questions about images requiring an understanding of vision, language and common sense knowledge to answer, is one of these new challenges where the new multimedia content representations allow common sense cross-modality analysis[252].

All these recent advances open new opportunities for developing systems to analyze and retrieve multimedia content, with direct applications in the media industry. The current focus of researchers in the field is to design new gigantic models called transformers as unifying models that receive all available modalities of multimedia content and provide solutions for many different tasks all at once (transformers are discussed in section 5.4). However, the design of such gigantic models is a challenge, even more when addressing multimodal data. An example of a deep transformer network for multimedia content is presented in Figure 62.

---

[246] G. E. Dahl, M. Ranzato, A. Mohamed, G. E. Hinton, Phone recognition with the mean covariance restricted Boltzmann machine. In NIPS. pp. 469-477, 2010.

[247] A. Krizhevsky, I. Sutskever, and G. E Hinton. Imagenet classification with deep convolutional neural networks. NeurIPS, 25:1097–1105, 2012.

[248] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

[249] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016.

[250] D. Tran, L.r Bourdev, R. Fergus, L. Torresani, M. Paluri, Learning spatiotemporal features with 3D convolutional networks. In ICCV, pages 4489-4497, 2015.

[251] R. Collobert and J. Weston. A Unified Architecture for Natural Language Processing: Deep Neural Networks with Multitask Learning. In ICML, 2008.

[252] Visual Question Answering (VQA): https://visualqa.org/

Until now, the advances in deep learning have not directly benefited the design of a new generation of search engines. Very few recent works investigate this possibility[253], even though the potential is clearly and unanimously identified.

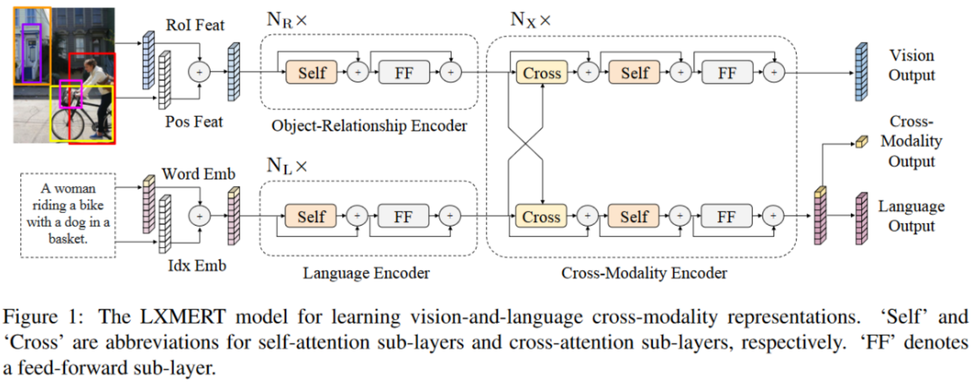


Figure 62: The LXMERT model for learning vision-and-language cross-modality representations[254].

Research challenges

As mentioned in the previous section, very few research works have been recently investigating the potential of integrating deep learning in designing a new generation of search engines, even though this has already been identified as a promising field of research and development.

The recent focus on transformers as central models dealing with all available modalities and addressing many different tasks all at once may be a game-changer. Let us imagine news producers searching audio-visual archives to support a news story with selected video, film writers searching film or script archives to get ideas, users searching the internet to find content that they like or need, music producers searching music with specific characteristics to match with film scenes or a textual story maybe matching some of the lyrics, game developers searching 3D content to find visual assets for a game level, etc.

Two serious concerns with regard to transformer architectures are the resources required both in terms of data to train the models and in terms of computation and energy consumption. This is true for textual data[255] but it will be even more challenging with these models applied to every possible modality and context.

There are several challenges to tackle in order to really leverage deep learning based search engines, or multimedia content retrieval. First, there is not yet any network architecture that reaches consensus to address all the aforementioned standard search contexts as it could be the case now for convolutional neural networks for image classification.[247,248,249] The research

[253] T. Teofili. Deep Learning for Search. In Manning Publications, June 2019.
[254] Image source: H. Tan and M. Bansal, LXMERT: Learning Cross-Modality Encoder Representations from Transformers, In EMNLP, 2019.
[255] Strubell, E., Ganesh, A., & McCallum, A. (2020). Energy and Policy Considerations for Modern Deep Learning Research. Proceedings of the AAAI Conference on Artificial Intelligence, 34(09), 13693-13696. https://doi.org/10.1609/aaai.v34i09.7123

works are still at their beginning both in academic labs and in companies (the most recent advances in transformers for natural language processing are led by companies such as OpenAI, Facebook, or Google).

The other challenge with regard to these gigantic models is related to the resources they require for training, as previously explained. One possible path to overcome this issue is to build hybrid models combining symbolic and non-symbolic approaches. This path is already explored, in particular for the visual question answering task, which can be addressed either as a multimodal retrieval task or as a cross-modality retrieval task, integrating external sources of knowledge.[256,257] Furthermore, using external knowledge may help to reduce the amount of required training samples. For instance, if a model can combine given grammatical rules with statistical analysis from textual data, less training data may be required.

Thus, the main research challenges are first to design and train sophisticated models which will require less annotations while combining several modalities, and second to integrate external knowledge to reduce the cost of building these models.

Societal and media industry drivers

Vignette: Multi-modal and cross-modal content search in vast multimedia data lakes

Chloé is a journalist searching a multimedia data lake for multimedia content that could be used for enriching her news story on the beginning of the French presidency of the EU. She is looking for content that could either be complementary sources of information on the main subject, or have a direct link with it, or at least be useful to illustrate the context of the story: she may look for a video of a historical discourse from the first French president of the Commission, Jacques Delors, or retrieve a sentence Jacques Delors has said on the importance of the EU during an international crisis (as it is currently the case with Ukraine and Russia). She may also look for pictures from the participation of President Emmanuel Macron in previous European Summits. She could also explore the recent discourses or recent interventions made by President Macron about the EU. Chloé will thus need several tools: first, a search engine very similar to the current search engines but with more sophisticated cross-modal capabilities, which would allow searching for images given a textual description of the queried content, providing an image to retrieve videos with similar visual content, or providing audio data to retrieve a corresponding video. A second tool would allow a multimodal analysis of the multimedia content to jointly exploit different modalities, hence providing richer content and a deeper analysis of this content. For instance, a sophisticated joint audio-visual speech recognition solution would allow to efficiently subtitle any conference or discourse and give access to its content by keyword search. This would allow Chloé to quickly find situations of tension between France and another European country in the history of the EU.

[256] K. Ye, M. Zhang and A. Kovashka. Breaking Shortcuts by Masking for Robust Visual Reasoning. WACV, January 2021.
[257] Q. Wu, P. Wang, C. Shen, A. Dick, A. van den Hengel. Ask Me Anything: Free-Form Visual Question Answering Based on Knowledge From External Sources. In IEEE CVPR, pp. 4622-4630, 2016.

To search for multimedia documents, Chloé has access to a professional multimedia data lake where content has been enriched with a lot of metadata (e.g. location and date of acquisition, topics present in the document, event related, media origin, etc.). It is actually easier to retrieve content using such metadata but many documents of the data lake have not been yet enriched with similar metadata. Furthermore, if Chloé wants to retrieve photos similar to the photos she provides to the system, metadata may not be the most relevant modality and a system directly working on the visual content will be more efficient. However, reasoning on metadata may help shortcut exploring the whole data lake and spending too much time on retrieving the best documents through different multimedia modalities. Thus, when Chloé finds new documents without metadata (e.g. a new photo, multimedia documents coming from another data lake, etc.), and she finds matching documents among the ones enriched with metadata, she can then propagate the metadata to the newly found documents (the location may be the same, the persons or the concepts pictured could be the same, etc.). Searching in multimedia content associated with metadata requires more sophisticated techniques combining reasoning and deep learning. This is a challenging new field for research.

Future trends for the media sector

Up to now, most of the contributions from private companies on new AI techniques to exploit multimedia content are made by big tech companies such as the GAFAM (Google-Amazon-Facebook- Apple-Microsoft) or the BATX (Baidu-Alibaba-Tencent-Xiaomi). Media companies are more usually customers of the solutions developed by Facebook and Google. Thus, media companies depend on the GAFAM and do not handle the processes at the core of the information retrieval system; they get what the algorithms designed by the GAFAM provide. It is urgent that media companies invest in this field to drive innovations more precisely towards their own specific needs and vision. Indeed, if the algorithms to retrieve information in huge multimedia lakes are biased, the search will not be accurate or even valid and may carry wrong information. Handling their own algorithms improves the possibility to control their weaknesses.

An efficient solution for audio-visual speech recognition, benefiting from both modalities, would allow to convert the speech in videos (interviews on TV, conferences, etc.) into a huge amount of textual content, a lot easier to search into and retrieve information from it. Another possible output of these multi-modal/cross-modal contexts, would be new information systems and search engines, able to jointly exploit prior knowledge and existing metadata with statistical models. By a simple text query a user could find/retrieve all relevant multimedia content no matter the available modality: a textual query could be matched with a video without sound; music producers searching music with specific characteristics (of rhythm and Minor tune) to match with film scenes; film writers searching film or script archives to get ideas.

Media and entertainment industry (news, film/TV, music, publishing, social media, games, etc.) would directly benefit from efficient content indexing and search. These companies are the ones producing the content but currently they need powerful intermediates to exploit that content. Being independent from external companies, which are also coming on the market of multimedia content producers, sounds both reasonable and also more efficient since no one better than the experts who produced the content could design a more accurate, fast and reliable information search engine.

This would also allow media companies to offer better services for the user to access a targeted content more accurately and fast, while preserving users' privacy. Working on designing specific systems to better exploit multimedia content would allow to automatise tedious tasks such as annotating new content, propagating new information associated to the multimedia content, confronting the search engine to the consistency of data. This would finally allow to enhance current multimedia content analysis workflows, and for each media company to more efficiently exploit and monetise their own content or UGC.

Goals for next 10 or 20 years

A significant goal is the emergence of new search engines for multimedia data that benefit from the advances in deep learning to solve multimodal and cross-modal queries, for instance to solve the well-known problem of visual question answering. Based on current progress in the field, such solutions could be available on the market within the next few years.

Other applications could also concern a better "dialog" between multimedia (i.e. unstructured content) and metadata (structured content) with the emergence of brand new approaches allowing a cross-fertilisation of these two information resources, resulting in enriched and improved content representation (in both unstructured and structured space). This phase will require more time than the previous one and some preliminary solutions may appear on the market within the next 5 to 10 years.

6.2 Media summarisation – The case for video

Contributors: *Evlampios Apostolidis (CERTH), Vasileios Mezaris (CERTH)*

Current status

Video summarisation technologies aim to create a short synopsis that conveys the important parts of the full-length video. In terms of presentation format, the produced summary can be either static, composed of a set of representative video frames (a.k.a. ***video storyboard***), or dynamic, created by stitching video's most important and informative fragments in chronological order to form a shorter video (a.k.a***. video skim***). One advantage of video skims over static sets of frames is the ability to include audio and motion elements that offer a more natural story narration and potentially enhance the expressiveness and the amount of information conveyed by the video summary. Furthermore, it is often more entertaining and interesting for the viewer to watch a skim rather than a slide show of frames. On the other hand, storyboards are not restricted by timing or synchronisation issues and, therefore, they offer more flexibility in terms of data organisation for browsing and navigation purposes.

During the last couple of decades, several attempts were made by the relevant research community to automate video summarisation. Currently, the focus is mainly put on methods that try to learn how to perform video summarisation by exploiting the learning capacity of deep network architectures. Most of these methods rely on datasets with ground-truth human-generated summaries (such as, SumMe[258] and TVSum[259]), based on which they try to discover the underlying criterion for video summarisation. However, the amount of currently-available data is relatively small, and the generation of ground-truth data (usually in the form of video summaries or annotations indicating the importance of video frames) is a time-consuming and tedious task.

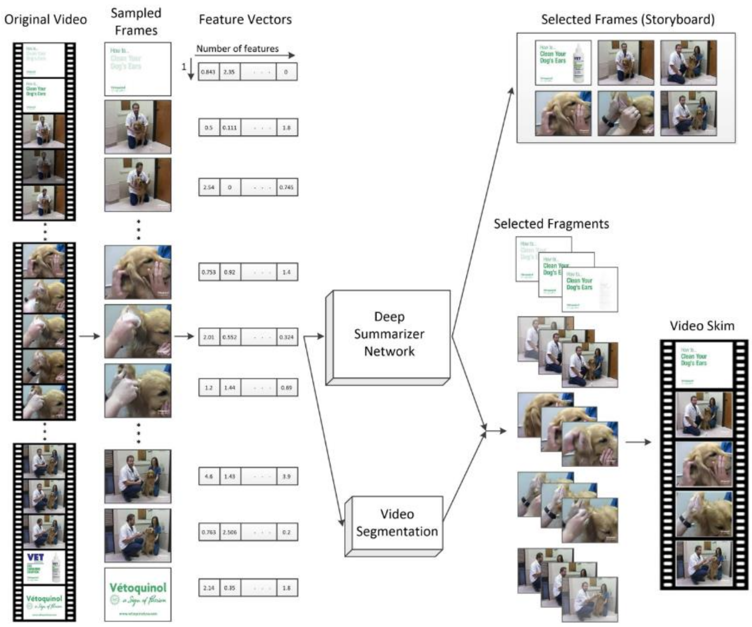


Figure 63: High-level representation of the analysis pipeline of deep-learning-based video summarisation methods for generating a video storyboard and a video skim.[260]

These limitations resulted in a constantly-increasing interest of the relevant research community, on the development of deep-learning-based approaches that can be trained without the use of extensively-annotated ground-truth data. A recent survey on deep-learning-based video summarisation methods[261], showed that ***unsupervised video summarisation*** methods can be highly-competitive compared to the best-performing supervised approaches. Moreover, the use of less-expensive weak labels - with the understanding that they are

[258] M. Gygli, H. Grabner, H. Riemenschneider, and L. Van Gool, “Creating Summaries from User Videos,” in European Conf. on Computer Vision (ECCV) 2014, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 505–520.
[259] Y. Song, J. Vallmitjana, A. Stent, and A. Jaimes, “TVSum: Summarizing web videos using titles,” in 2015 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), June 2015, pp. 5179–5187.
[260] Figure provided by the authors. Source: DOI:10.1109/JPROC.2021.3117472.
[261] E. Apostolidis, E. Adamantidou, A. Metsai, V. Mezaris, I. Patras, "Video Summarization Using Deep Neural Networks: A Survey", Proceedings of the IEEE, vol. 109, no. 11, pp. 1838-1863, Nov. 2021. DOI:10.1109/JPROC.2021.3117472.

imperfect compared to a full set of human annotations - can be another option to build good summarisation models. Figure 63 above presents the typical analysis pipeline of deep-learning-based video summarisation methods.

Despite the fact that some video summarisation methods already exhibit good performance according to the established evaluation protocols and datasets, nowadays the practical use of integrated tools and applications for video summarisation is at a very early stage and quite far from being a common procedure of the media management workflow. In most cases, the preparation of a video summary requires the observation of the full-length video and the manual selection of the most suitable pieces of video content by a human expert; and depending on the length of the video this procedure can be a really time-demanding one. Taking into account the number of video material created or obtained on a daily basis by the stakeholders of the media sector, the production of video summaries could require considerable amounts of human and time resources, and possible shortages in these resources could significantly harm the revenue generation potential from the use of these video materials.

This problem is further emphasised by the growing use of different communication channels with varying requirements (such as social networks and video sharing platforms) by media organisations that necessitates the production of ***different summaries for the same piece of video content***. These requirements are usually related to the needs of the targeted audience; for example Twitter users are used to get very short videos, the users of Facebook are familiar with a bit longer videos, while the users of the YouTube platform can spend even more time when watching a video. All the above point out a serious lack of mature video summarisation technologies, that could significantly assist professionals by facilitating and, most importantly, accelerating video summary production.

A paradigm of an integrated tool which can assist the production of video summaries that are tailored to the specifications of different communication channels, is the "On-line Video Summarisation Service"[262] (Figure 64). This tool harnesses the power of artificial intelligence[263] to automatically generate video summaries. It takes as input a video and produces different versions of a video summary with adapted length and format for publication on different social media platforms. Based on its functionality, the user can accelerate the production of engaging video summaries for multiple on-line audiences.

[262] C. Collyda, K. Apostolidis, E. Apostolidis, E. Adamantidou, A. Metsai, V. Mezaris, "A Web Service for Video Summarization", Proc. ACM Int. Conf. on Interactive Media Experiences, Barcelona, Spain, June 2020 DOI:10.1145/3391614.3399391. Online demo available at http://multimedia2.iti.gr/videosummarization/service/start.html

[263] E. Apostolidis, E. Adamantidou, A. Metsai, V. Mezaris, I. Patras, "AC-SUM-GAN: Connecting Actor-Critic and Generative Adversarial Networks for Unsupervised Video Summarization", IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 8, pp. 3278-3292, Aug. 2021. DOI:10.1109/TCSVT.2020.3037883

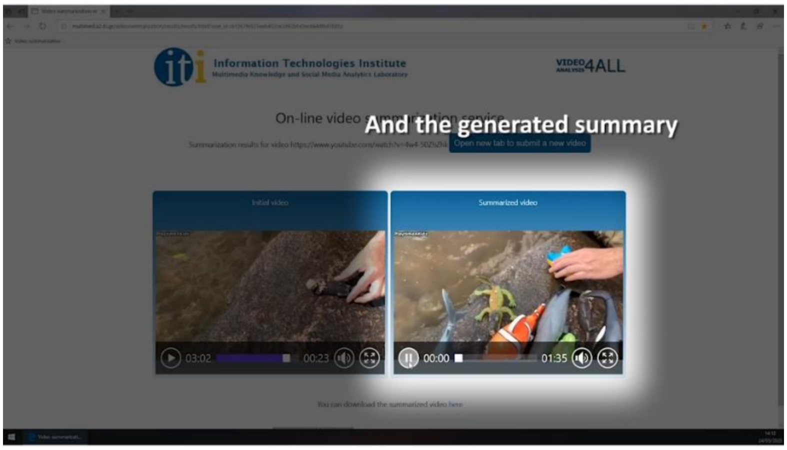


Figure 64: An online video summarisation service [264].

Research challenges

To identify the research challenges in the field of video summarisation, one should consider the current state of the art in terms of summarisation methods, and the current (and possibly future) requirements of the media sector for video content adaptation and re-purposing. With respect to the former, the major research direction is towards the development of supervised algorithms. However, there is an ongoing and increasing interest in the design and development of ***unsupervised video summarisation methods***, mainly propelled by the ***limited amount of training data***. With regards to the latter, the traditional communication channels of media organisations (e.g., TV streams, webpages and online archives) have been framed by the adoption and wide use of modern communication instruments that include social networks and video sharing platforms. Nevertheless, each one of these new instruments is associated with different (mainly audience-driven) requirements about the ***format of the distributed content***. Hence, experts working in the media sector should deal with the preparation of different adapted versions of a given piece of video, in order to reach different audiences and increase the potential for effective media re-use.

Given the above described landscape, we believe that a core research challenge is the development of effective video summarisation methods that can be trained without or using limited supervision. In this way, the research community will be able to tackle issues associated with the restricted amount of annotated data, and to significantly diminish (or even completely eliminate) the need for laborious and time-demanding data annotation tasks. Moreover, such developments can be a game changer considering the need to train different models for each different type of video, as they can provide a solution for building powerful summarisation models that can be easily adapted to the requirements of different application domains (e.g. targeting the production of movie trailers, or the generation of videos with the highlights of a sports event).

[264] Screenshot taken from: YouTube, CERTH ITI Video Summarization Service v1.0 (March 2020) https://www.youtube.com/watch?v=LbjPLJzeNII

With respect to the development of ***fully-unsupervised video summarisation methods***, most of the existing approaches learn summarisation by trying to increase the representativeness of the generated summary with the help of summary-to-video reconstruction mechanisms[265]. So, the main criterion for building the summary is the coverage of the full-length video. The challenge for the relevant research community is to identify additional ***criteria for selecting the key-parts of the video***, and to formulate these criteria in ways that enable their integration into the unsupervised learning process. Such criteria can be associated, for example, with the diversity of the visual content of the summary (to avoid the existence of similar and possibly redundant information), its temporal coherence (to provide a meaningful and appealing presentation of the video story), or its alignment with the core semantics of the video (to focus on parts of the video that are associated with the video's topic or subject).

With regards to the development of ***weakly-supervised video summarisation methods***, the challenge is to discover effective ways that allow editors to intervene in the summary production process, so that the produced video summary is aligned with user-specified rules and requirements. By taking into account user-profile data (e.g., indicating a user's preferences), the research community could provide solutions that facilitate the provision of personalised video summaries. Another, more aspiring scenario would involve the use of an on-line interaction channel between the user/editor and the trainable summariser. So, the challenge here is to build solutions that combine video summarisation and active learning algorithms, in order to incorporate the ***user's/editor's feedback*** with respect to the generated summary[266]. Such developments will be extremely useful for the practical application of summarisation technologies in the media sector, where complete automation that diminishes editorial control over the generated summaries is not always preferred.

Given the plethora of online-available examples of video summarisation, another research challenge could be the design of an effective methodology for ***learning from unpaired data*** (i.e., using raw videos and video summaries with no correspondence between them)[267]. In this way, weak supervision is associated to the collection of the appropriate sets of data based on the targeted application domain (e.g., unpaired groups of movies and movie trailers). Such a data-driven weak-supervision approach would eliminate the need for defining hand-crafted functions that model the domain rules (which in most cases are really hard to obtain), and would allow a deep learning architecture to automatically learn a mapping function between the raw videos and the summaries in the targeted domain.

Last but not least, researchers working in the field of video summarisation should target the development of mechanisms that ***remove bias and provide human-interpretable explanations***

---

[265] E. Apostolidis, E. Adamantidou, A. Metsai, V. Mezaris, I. Patras, "Unsupervised Video Summarization via Attention-Driven Adversarial Learning", Proc. 26th Int. Conf. on Multimedia Modeling (MMM2020), Daejeon, Korea, Springer LNCS vol. 11961, pp. 492-504, Jan. 2020. https://doi.org/10.1007/978-3-030-37731-1_40
[266] A. G. del Molino, X. Boix, J. Lim, and A. Tan, "Active Video Summarization: Customized Summaries via On-line Interaction," in Proc. of the 2017 AAAI Conf. on Artificial Intelligence. AAAI Press, 2017.
[267] M. Rochan and Y. Wang, "Video Summarization by Learning From Unpaired Data," in 2019 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR), June 2019, pp. 7894–7903.

about the decisions made by an AI-based video summarisation method. In this way, the provided solutions for automated video summarisation will offer the needed transparency to the end-users, increasing in this way the level of trust between machines and humans, and improving user experience.

Overall, the modern deep learning architectures have already shown their great potential to learning the main principles of generic video summarisation. From now on, the goal for the research community is to effectively tackle the aforementioned challenges, and push the barriers for making these architectures easily adaptive to the video summarisation needs of several domains and application scenarios. In this way, building on existing technologies with demonstrated content adaptation and re-purposing capabilities[268], it will provide mature solutions that meet the requirements of the media industry, and highly accelerate the video content adaptation and distribution tasks of media professionals.

Societal and media industry drivers

Vignette: Creating teasers, trailers and video summaries for the promotion of TV shows in TV and social media

Jane is a media and social media professional with an expertise on the preparation of highly-engaging video content. Over the last five years, she has been working in a large TV network with an active presence in social media networks (Facebook and Twitter) and video sharing platforms (YouTube). As part of her daily job, Jane deals with the production of different types of summaries for the episodes of five popular TV series. More specifically, a complete video summary is created in order to be added in the beginning of each episode and provide a synopsis of the previous one; a shorter video trailer is produced to advertise each new episode via the TV program, the YouTube channel and the Facebook account; and a very shot teaser video is generated to promote the release of new episodes on Twitter. A couple of years ago, this procedure used to be quite laborious and time-consuming. For every single video, Jane had to watch the entire content, spot the most appropriate parts of it, and then produce summaries of different length by deciding which of these parts will be included in each different type of summary. Usually, she was starting by producing the longest and more complete one that was used during the production of the next episode. Then, she was preparing the shorter trailer video for the TV program, the YouTube channel and the Facebook account. Finally, she was working on the creation of the short teaser video that would be posted on Twitter. In total, for a 45-minute episode, Jane spent approximately three hours to prepare all the different summaries.

However, the last two years, Jane's work has been significantly accelerated, as Elena, the director of the Media Management Dept., decided to buy a tool for video summarisation that is based on AI technologies. Using this tool, Jane is now able to analyse an episode and quickly get recommendations about the most informative and story-telling parts of it. Based on the

[268] L. Nixon, K. Apostolidis, E. Apostolidis, D. Galanopoulos, V. Mezaris, B. Philipp, R. Bocyte, "Content Wizard: demo of a trans-vector digital video publication tool", Proc. ACM Int. Conf. on Interactive Media Experiences (IMX), June 2021. DOI:10.1145/3452918.3468083.

automatically provided explanations about the recommended pieces of information, she can easily decide whether she needs to check other parts of the video as well. If there is such a need, then using the generated storyboard of the episode (i.e., a static visual summary composed of a set of representative frames of the video) Jane is able to quickly inspect other parts of the video and replace some of the recommended ones, by other, manually selected by her. Having this first version of the video summary (i.e., the most complete one) available for the production of the next episode, Jane can then create shorter versions of this summary, by simply adjusting a parameter in the interface of this tool, that is associated with the summary duration. So, after a few clicks and editing checks, she can create the trailer and the teaser of the episode. Moreover, for each type of generated summary, she can immediately create different versions of the video file (with different size and bit-rate) that are compatible with various devices (laptop, table, smartphone) and networks (5G, Wi-Fi, networks of limited bandwidth). Based on the process described above, the production of all these different types of summaries for a 45-minute episode now takes approx. 30 minutes. So, using the AI-based video summarisation tool, Jane needs significantly less time to produce effectively-customised and highly-engaging video summaries; and after spending some of the saved time on the design of the social media campaigns for these five TV series, she achieved much higher audience engagement!

Future trends for the media industry

The development of mature AI-based technologies for video summarisation will be a game changer in the media industry. In the following, we discuss some cases that highlight how these technologies will be used as part of the data analysis workflows of media organisations to accelerate video content adaptation, re-purposing and re-use:

- Process different types of already existing proprietary video data (including both full-length videos and their summarised versions) with powerful AI-based video summarisation tools that can automatically identify the main summarisation patterns for each different type of video content.
- Utilise the trained version of these tools to analyze a new video and quickly get recommendations about the key parts of it that should be taken into account when producing the video summary, and check the automatically provided explanations about the tools' choices with respect to the key parts of the video.
- Constantly improve the performance of the AI-based video summarisation workflow by providing feedback about the provided recommendations about the key parts of the video, based on the ability of the video summarisation tools to actively learn from and adapt to the user's preferences.
- Produce different summarised versions of the same video according to different criteria about the content and duration of the summary, thus accelerating the production of e.g., teasers, trailers and summaries for an episode of a TV series.
- Create different versions of the produced summaries for distribution and consumption via different communication channels (e.g., TV, web-TV, account in social media networks, channels in video sharing platforms) and different devices (e.g., smart-TVs, laptops, tablets, smartphones).

- Enhance the AI-based video summarisation workflow by integrating the available profile data about the viewers/subscribers of the different on-line channels of the media organisation, to offer highly-personalised summaries that match each viewer's interests; e.g., given a 2-hour video of a TV show about traveling to different cities, generate a summary for food lovers, that focuses on the visits at the city restaurants, and another summary for people interested in nature, that shows scenes from rivers and mountains near the city

Goals for next 10 or 20 years

With respect to the goals for the next one or two decades, we foresee that: i) ***repositories of well-trained models*** for the most common video summarisation scenarios (e.g., targeting the production of summaries for news shows, documentaries, sitcoms and the generation of movie trailers and teasers) will be available; ii) AI video summarisation systems will be able to ***discover more complex, domain-specific rules***, and facilitate, for example, the production of short videos with the highlights of sports, music or other events; iii) AI video summarisation systems will be quickly ***adapted to new types of data*** (e.g., medical videos, educational videos, videos captured from surveillance cameras) based on the transfer learning mechanisms of the integrated tools; iv) AI video summarisation systems will be capable of ***summarising more complex types of video content, such as 360-degree video, AR video and XR video***; and v) ***highly compact versions of powerful AI video summarisation methods*** will be available for use in mobile devices, allowing the instant generation of summaries for any recorded video; thus enabling new forms of user-generated content and empowering citizen journalism.

6.3 Automatic multimedia content creation

Contributors: *Lorenzo Seidenari (UNIFI), Marco Bertini (UNIFI)*

Current status

Content creation is the pillar of the multimedia industry. Media companies, broadcasters, artists, and media professionals in general all make a living out of producing a large stream of multimedia data. While most of the professional content is still manually edited and originated, several computational methodologies have risen in support of content creators (e.g. see section 5.2). Apart from AI based methods to support editing, which have been covered in section 6.2, a lot of effort has been made by researchers to enable machines to produce novel unseen content by directly learning from data (Figure 65).

Figure 65: Examples of faces generated using StyleGAN algorithm*[269]*.

Generative art lays its foundation on methods that optimise large deep network models to create outputs that are reasonable with respect to a set of training data. Amazing results have been shown by methods like DALLE[270] and VQGAN+CLIP[271]. Both approaches leverage large image and language datasets. While the former usually yields precise depiction of the provided sentence, the latter has mostly been used to obtain oneiric imagery. DALLE's main strength is also its main drawback: a large 12 billion parameter model that requires 250 million image-text pairs mined from the internet. Considering current video standards (4K), one major showstopper for these methods is the high computational demand and their limited output resolution. Direct generation of video is also sought[272], which of course is affected even more by the above challenge.

A large amount of content is continuously produced and streamed online. On top of this, a bulk portion of existing digital content has largely been acquired and produced over the last decades with a plethora of not always well-performing formats. Modern video consumers are used to high quality, high resolution devices. Current broadcasters, especially on pay-per-view, aim at delivering 4K content. Media editors and producers often struggle to reuse content and keep the overall product quality high. Several phenomena intervene in degrading clips. A video may be acquired digitally at a low resolution with some older formats (e.g. MPEG2), some content may not be available in colour or at the desired final resolution.

Content enhancement is the task of improving quality of media from a possibly low quality or corrupted source. Recent TVs implement AI based enhancement algorithms directly in hardware. On the research end, we witness a rush towards deep learning based methods capable of blindly reconstructing missing information either by correcting defects or artefacts or by filling in lacking pixels when upsampling. First attempts at improving images and video came from classical convolutional neural networks trained to translate a low quality input image

[269] Images generated randomly by the Random Face Generator (This Person Does Not Exist) at https://this-person-does-not-exist.com/en

[270] Ramesh, Aditya, et al. "Zero-shot text-to-image generation." International Conference on Machine Learning. PMLR, 2021

[271] VQGAN+CLIP Notebook: https://colab.research.google.com/drive/1ZAus_gn2RhTZWzOWUpPERNC0Q8OhZRTZ

[272] VIDEO-GPT: https://wilson1yan.github.io/videogpt/index.html

into a higher quality one[273]. More recent work leveraged techniques based on generative adversarial networks trained in a conditional fashion[274] (Figure 66). Since many distortions may be present in images, even at the same time, some effort has also been made to make a universal model[275], capable of restoring images blindly from unknown distortions.

(a) (b) (c) (d)

Figure 66: Effect of detail enhancement on a compressed picture (a). In (b) a clear degradation is shown. Standard non AI based methods are not able to improve the result (c). Finally, a GAN based enhancer shows a pleasant detail (d)[274].

Sometimes mixing black&white video with colour sources may be a style choice also guided by the will of the narrator to make the viewer focus on the change of a historic period. In some scenarios, again guided by a different artistic need, colour shall be recovered or in general improved so that the final product is consistent. A preliminary solution to this challenging problem is a technique called colourisation. In colourisation, deep networks are trained to translate images from black&white to full colour. In this task, networks shall preserve shape but add colour information pixelwise. Issues usually regard the temporal consistency of such solutions and the need of preserving semantic information that could be lost when colour is transferred wrongly.

Research challenges

The main challenge for current enhancement methods is **dealing with multiple sources of disturbance** in the input media to be improved. Since the process of degradation is non additive, meaning that a processing pipeline may involve a sequence of compression, rescaling and other kind of noises that are hard to remove individually or even detect. As an example, a video digitised from a partially compromised physical media with low quality digital coding may be hard to restore with current methods.

[273] Svoboda, Pavel, et al. "Compression artifacts removal using convolutional neural networks." arXiv preprint arXiv:1605.00366 (2016).
[274] Galteri, Leonardo, et al. "Deep universal generative adversarial compression artifact removal." IEEE Transactions on Multimedia 21.8 (2019): 2131-2145.
[275] Wang, Xintao, et al. "Real-esrgan: Training real-world blind super-resolution with pure synthetic data." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021

In respect to approaches for novel content generation, the current main limitation lies in the capability to **scale in terms of temporal and spatial resolution**. Current face generators cannot go above 1MP resolution and video generators are far behind. Methods able to create high-quality high-resolution complex scenes are not available yet.

Another challenge is that of **bias in training data**. As every data driven approach, especially when dealing with large deep learning models the issue with dataset bias can lead to extreme failures. As a very well-known example has recently shown, the lack of diversity in face datasets may significantly alter the output of enhanced faces to the end that most of the visual attributes are corrupted (see Figure 67). The ethical ramifications of such failure modes are severe and must be addressed.

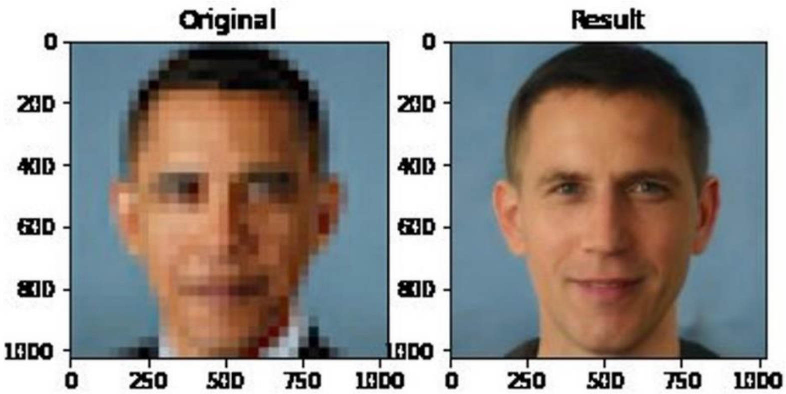


Figure 67: Example of the PULSE[276] algorithm applied to a pixelated face of Barack Obama. The algorithm reconstructs the face as that of a white man[277].

Societal and media industry drivers

Vignette: Automating video production via reuse of content and original content creation

Francesca is a video editor that works in a major broadcasting company. The company broadcasts content of all sorts from documentaries to live sport events. Francesca starts her productions designing a digital storyboard to show how she wants to visually portray the programmes before they are produced. The production and selection of the visual materials of the storyboard sometimes is slow, since she can't find archive material that is similar enough to the description of the planned scripts and thus, she must draw or produce this storyboard content with the help of a 3D artist, resulting in a slower production or, sometimes, even in misunderstandings with the production team.

Moreover, Francesca is often tasked with the editing of clips from a diverse set of sources, especially when dealing with news reports and documentaries. To keep the final product quality high, since the editing process often has to deal with intertwining different qualities and

[276] Menon, Damian, et al. "PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models", Proc. of CVPR 2020.
[277] Image source: Twitter (@Chicken3gg) - https://twitter.com/Chicken3gg/status/1274314622447820801

resolutions of media, Francesca often resorts to hand-tuning the appearance of clips using video editing software. As a video editing technician, Francesca is often consulted regarding the technical process involved during live streaming events. In this situation, a bulk of streams from a very diverse set of cameras with different resolutions, framerate, format and colour calibration parameters are pooled together to be mixed in real time. Event producers are worried that the continuous shift in appearance hurts the end user experience and enrol Francesca in the process of adjusting, in a best effort manner, the appearance of all streams in a way that the final product has the highest possible quality. Unfortunately, due to many factors such as weather conditions that may affect lighting, equipment variation and also bandwidth limitation this effort is continuous and sometimes the sought result is far from the actual final product.

Recently Francesca's company established a tight collaboration with a couple of start-ups: one is a company that provides a set of tools to create stock footage materials and the other provides software for image and video enhancement. Together they started a proof-of-concept that led to a set of tools and systems that allow to automatise all processes related to video editing. Thanks to this collaboration, the companies grew and made the PoC results into actual industrial products that revolutionised the editing business. Now Francesca can rely on such products to improve the final quality of archival content prior to video production and has sped up the production of new programs, reducing the planning and design phase. Using the tools for stock footage generation, she can produce thumbnails that match the descriptions in the scripts; the same tools can produce new animated backgrounds and 3D graphics footage, thus reducing production costs. Using the tools for video enhancement, live streams improved in quality and the broadcaster widened its 4K offer increasing their subscriptions. Thanks to AI based automatic content enhancement, Francesca can spend more time following the artistic and technical process of video editing while content is processed offline. The lack of need of manual tuning makes it possible to focus on content selection. Regarding live-streamed events, producers are happier since they have less uncertainty on video quality when switching from one feed to another.

Future trends for the media sector

The automatic creation of content has a major drive in the continuous request for high quality media products. We highlight some trends that we believe will lead the development of new AI based media product:

- Developing **robust and efficient video enhancement services** allowing to edit content in real time and increase the throughput of video professionals.
- Deployment of **content enhancement AI on TV hardware** allowing end users to benefit from the maximum quality possible, independently from their connectivity and the broadcasted signal quality.
- **Repurposing and revamping of archive materials** allows to amortise the production costs by increasing the reuse; this requires adapting the low level visual features of recorded material (e.g. colour correction, colourisation, super resolution) for a new production. Automatising the process reduces costs and facilitates reuse.

- Instead of restoration, archive materials could be used as “seeds” for the **generation of new content adapting visual styles**, e.g., from movies to cartoons or 3D graphics and vice-versa.
- Integration of **automatic original content creation** into existing content production tools. Currently some softwares are already exploiting AI to provide smart layouts for presentations or to fix image appearance without user intervention. Current image generator AIs could replace the role of an **image search engine**. Instead of looking up for the best stock photo or the best clipart to integrate into a digital product, such tool could just generate a set of original images based on the query of a user.
- **Creation of deepfakes**, either as persons (full bodies or talking heads) or as settings and scenes, could lower production costs allowing also smaller companies to produce a larger variety of products, e.g., those that require expensive film sets.
- The videogame industry could benefit from **automated creation of graphical content and assets**, like backgrounds, textures, cut scenes, **or audio content**.

Goals for next 10 or 20 years

Rapid progress is expected in this field in the next 10 years. Below we summarise some goals.

One goal is the ability to **disentangle internal representations** in order to not just create realistic images and video but also to fully control the appearance via semantic and geometric attributes. This means that generated art or content should not just be pleasant and realistic but a user should be able to edit specific properties such as the arrangement of objects, their colour or texture without corrupting the overall appearance of the generated example.

Regarding content enhancement tools, we expect models to be able to **deal with multiple sources of degradation at the same** time blindly. Like current image classifiers being resilient to changes in image resolution, object scales and appearances, the same should be sought for content enhancement solutions.

Currently, the correspondence between low/high quality media is 1-to-1, while in the future, we expect to have methods able to **capture the signal from a few examples** and exploit this source to improve a larger amount of data. Currently, this is possible for video call like videos such where once every few seconds a high-quality face is transmitted and the receiver can enhance the current low quality stream incorporating information from high quality frames. This approach will likely be generalised to more complex types of videos, allowing for example to improve the quality and resolution of archival videos with few key high quality digital shots.

Finally, while current learning paradigms expect to work with paired media from different qualities, in the future **unpaired learning** will be likely possible allowing to leverage a larger amount of training data. In this scenario, enhancement tools will not just learn to invert some simulated degradation but given two large media sets with a source quality and a desired target quality will learn to restore content even without a direct pairing.

6.4 Affective analysis

***Contributors**: Ioannis Patras (QMUL), Niki Foteinopoulou (QMUL), Ioannis Maniadis Metaxas (QMUL), Ioanna Ntinou (QMUL), James Oldfield (QMUL), Christos Tzelepis (QMUL), Georgios Zoumpourlis (QMUL)*

Current status

Over the last two decades, there has been an increasing interest towards modelling human affect through the field of ***Affective Computing***. Affective Computing is *computing that relates to, arises from or deliberately influences emotion or other affective phenomena*[278]. One of the incentives of such research is the creation of ***empathetic machines***, i.e. machines that understand and interpret a human's emotional state and adopt their behaviour to give responses corresponding to our emotions and moods. Human affective and cognitive mental states are pivotal to human experience, and hence, the creation of empathetic machines can possibly influence mental health care through automatic depression detection[279], pain estimation[280] or post-traumatic stress disorder identification[281]. Other uses include automotive industry[282], education[283,284] and media as it will be discussed below.

Figure 68: Face of different identities and expressions.[285]

[278] R. Picard, "Affective Computing" MIT Technical Report #321 (Abstract), 1995
[279] J. M. Girard, J. F. Cohn, M. H. Mahoor, S. Mavadati, and D. P. Rosenwald. Social risk and depression: Evidence from manual and automatic facial expression analysis. In FG, 2013
[280] Xin X, Lin X, Yang S, Zheng X. Pain intensity estimation based on a spatial transformation and attention CNN. In PLoS One, 2021
[281] G. Stratou, S. Scherer, J. Gratch, and L.-P. Morency. Automatic nonverbal behavior indicators of depression and ptsd: Exploring gender differences. In ACII, 2013.
[282] C. Busso and J. J. Jain. Advances in Multimodal Tracking of Driver Distraction. In Digital Signal Processing for in Vehicle Systems and Safety, pages 253–270. 2012
[283] B. McDaniel, S. D'Mello, B. King, P. Chipman, K. Tapp, and a. Graesser. Facial Features for Affective State Detection in Learning Environments. 29th Annual meeting of the cognitive science society, pages 467–472, 2007
[284] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker. Multi-PIE. IVC, 2010
[285] Figure taken from: O. Ignatyeva, D. Sokolov, O. Lukashenko, A. Shalakitskaia, S. Denef, T. Samsonowa, "Business Models for Emerging Technologies: The Case of Affective Computing", In 2019 8th International Conference on Affective Computing and Intelligent Interaction Workshops and Demos (ACIIW) 2019 Sep 3 (pp. 350-355). IEEE.

Key for all of the aforementioned applications is reliable estimation of human affect. Most of the existing works on ***automatic emotion prediction*** rely on two psychological theories that can generally be distinguished according to the way emotion is modelled: to be categorical or continuous. In the *continuous approach*, which is coined ***Valence-Arousal-Dominance (VAD) Model***[286], emotions are described based on three dimensions in the range [-1,1], with *Valence* referring to the level of positiveness (-1 being negative and +1 being positive), *Arousa*l referring to the level of activation (-1 Being deactivated And +1 being activated), and *Dominance* referring to the level of control a person feels over a given situation (-1 being submissive and +1 being in-control (Figure 69). On the other hand, the *categorical approach* refers to ***seven basic emotions***, defined by Paul Ekman[287], namely Anger, Disgust, Fear, Happiness, Sadness, Surprise, and Neutral. Additionally, Ekman and Friesen[288] proposed another system for emotion analysis, which is based on the exclusive analysis of facial expressions, namely ***Facial Action Coding System***. This system defines emotional states as combinations of atomic facial muscle movements, coined *Action Units (AUs)*.

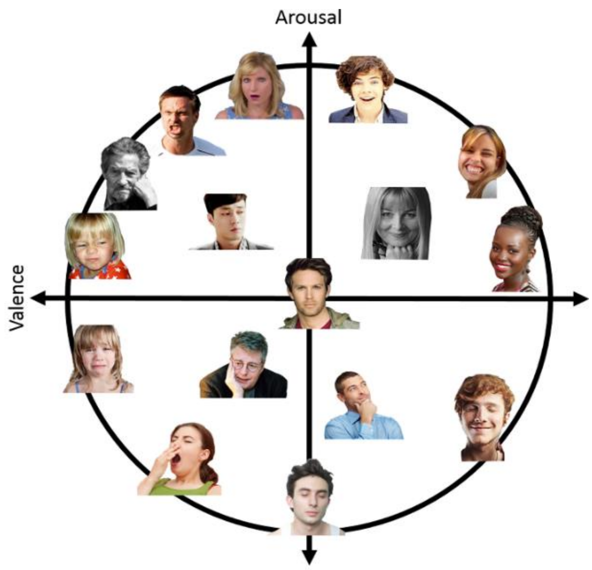


Figure 69: Illustration of Russell's dimensional emotion modelling scheme (Arousal-Valence).[289]

A significant body of work on emotion prediction is on unimodal affect prediction and in particular on facial expression analysis. This happens because facial signals are among the most important channels of nonverbal communication, thus offering discriminative cues for affective prediction. Based on this, there has been a significant development of datasets on the task of

[286] A. Mehrabian, "Framework for a comprehensive description and measurement of emotional states." Genetic, social, and general psychology monographs, 1995.
[287] P. Ekman and W. V. Friesen, "Constants across cultures in the face and emotion." Journal of personality and social psychology, vol. 17, no. 2, p. 124, 1971
[288] E. Friesen and P. Ekman, "Facial action coding system: a technique for the measurement of facial movement," Palo Alto, 1978
[289] Figure taken from: A. Mollahosseini, B. Hasani, M. H. Mahoor, "Affectnet: A database for facial expression, valence, and arousal computing in the wild", IEEE Transactions on Affective Computing, 2017 Aug 21;10(1):18-31

facial expression analysis with extensive Action Unit and VAD annotations. Other single modalities explored in the literature include voice and speech expressions[290], body gestures[291], gait[292], and physiological signals such as respiratory and heart cues[293], and more recently general context and scene background[294].

Although most of the works in affect prediction are focused on facial expression analysis, making predictions from a single modality can be challenging, especially because single sensory observations are ambiguous or incomplete. Indeed, many researchers advocate towards ***Multimodal Emotion Recognition***, i.e. the fusion of multiple modalities, based on two main arguments: a) *Accuracy increase*: different modalities can complement each other resulting in more accurate inference and, b) *Reliability of predictions*: single sensor observations can be corrupted leading to ambiguous, noisy or incomplete data. Hence, combining observations from multiple sources can potentially mitigate such challenges. To combine different modalities three different approaches have been investigated: feature-level (early) fusion, decision-level (late) fusion and model-level fusion. The main objective in such approaches is figuring out *when,* i.e. which stage of the training process, and *how,* i.e. with what weights, different modalities can be fused to leverage information from the most reliable ones. Overall, multimodal recognition has significantly advanced the task of emotion prediction in the wild in datasets like CMU-MOSEI[295] and IEMOCAP[296].

Research challenges

Despite the significant advances of affective computing in the last two decades, there are still numerous challenges that need to be addressed in the future. A number of them is discussed below.

Affect modelling across cultures. One recurrent debate in affective computing is whether emotion expression is a universal, biologically based construct or a social construct, i.e. affect expression changes across cultures. Paul Ekman argued that emotion is biologically defined, and hence, emotions are experienced and interpreted similarly across cultures. Indeed, this theory proposes six basic emotions (e.g. happiness, sadness, disgust) as the basic feelings someone can experience. Contrary to the basic emotions theories of Ekman, Russell and Barrett suggest that emotions change across cultures and proposed that emotions can be defined by three-independent dimensions: pleasant-unpleasant, tension-relaxation, and excitation-calm, i.e. the

---

[290] K. Scherer, T. Johnstone, and G. Klasmeyer. Vocal expression of emotion. Handbook of affective sciences, pages 433–456, 2003

[291] C. Navarretta. Individuality in communicative bodily behaviours. In Cognitive Behavioural Systems, pages 417–423. Springer, 2012

[292] T. Randhavane, A.t Bera, K. Kapsaskis, U. Bhattacharya, K. Gray, and D. Manocha. Identifying emotions from walking using affective and deep features. arXiv preprint arXiv:1906.11884, 2019

[293] B. Knapp, J. Kim, and E. Andre.´ Physiological signals and their use in augmenting emotion recognition for human–machine interaction. In Emotion oriented systems, pages 133–159. Springer, 2011

[294] R. Kosti, J.M. Álvarez, A. Recasens and A. Lapedriza, "Context based emotion recognition using EMOTIC dataset", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), 2019

[295] A. Zadeh, P.P. Liang, S. Poria, E. Cambria and L.P. Morency. “Multimodal Language Analysis in the Wild: CMU-MOSEI Dataset and Interpretable Dynamic Fusion Graph.” ACL (2018): 10.18653/v1/P18-1208

[296] The Interactive Emotional Dyadic Motion Capture (IEMOCAP) Database: https://sail.usc.edu/iemocap

Valence-Arousal-Dominance model. In the field of affective computing, some datasets are annotated with Ekman's models (6 basic emotions or AUs), others in the Valence-Arousal-Dominance model and others in a less structured manner with emotional words. There is a challenge in developing computational models that can learn with all of those annotations and learn mappings between them that are informed by theories in the field of psychology – those will be potentially conditioned on cultural and other contextual factors.

Dataset bias and inconsistent annotations. A shared problem across several domains is the fact that data collection conditions are inevitably limited, and hence the available training and testing datasets are biased. This is evident during cross-dataset experiments where we notice a significant discrepancy in performance when an algorithm trained on a dataset generalises well on the test set of the same data but performs poorly on a different dataset. This problem is more prominent in the field of affective computing, as emotion annotations can be very subjective thus can vary a lot between different annotators. This means that even if annotations are consistent within a specific dataset, it will be very hard to be consistent across different data. Given this, merging multiple datasets for affect prediction is a challenging task.

Class imbalance. One common challenge in datasets that use categorical emotions is class imbalance. This is traced back in the data acquisition process where collecting positive feelings is much easier than collecting data that correspond to negative feelings like anger or disgust. Overall, capturing or collecting data associated with negative emotions is a challenging task that needs to be addressed.

Design choices in multimodal affect recognition. While combining data from different sensors has advanced the field of affective computing, it comes with new challenges. To design a multimodal system, we first need to decide which modalities can be combined and how. Some modalities are co-occurring while others are not. For example, IEMOCAP and CMU-MOSEI datasets hold a similar set of modalities (facial expressions with the corresponding text and audio) and are typically employed together in the literature. Additionally, one critical design challenge is the way different modalities can be combined together, typically coined fusion. As stated above, there are three main fusion techniques, feature-level (early) fusion, decision-level (late) fusion and model-level fusion; however, it is not clear how modalities can be effectively fused given that there is inherent asynchronicity in the different data streams, interactions at potentially large time frames, and lack of large-scale data that would allow examining many different architectural choices. A challenge in this area is to design architectures that model the interplay between the different modalities at the appropriate level.

Context modelling. When trying to estimate the emotions that visual stimuli (e.g. images) elicit to human subjects, parsing the entire image and directly performing inference, is an approach with inherent limitations. Not all image regions contribute equally to emotion elicitation processing, and also individual objects/scene parts might not contribute alone to emotions, but through their interactions/relationships with other objects/scene parts. Stronger insights can be obtained through machine learning models that are capable of reasoning about object-to-object and object-to-scene interactions[297]. A first step towards this direction, is the development of techniques that achieve spatial localisation of regions that influence sentiment evoking, producing soft sentiment maps[298]. Also, the exploration of architectures that can learn emotion-related representations, taking into consideration the context around image regions[299]. Finally, the currently existing affective datasets do not suffice to develop robust affective models with capabilities as those described above. Some of the problems that the existing datasets have, are the following: i) small sample cardinality, ii) lack of fine-grained emotion annotations, iii) lack of multi-label emotion annotations, and iv) lack of spatially localised annotations. Thus, the curation of affective datasets with such content and annotations, will help to build stronger machine learning models.

Causality. In scenarios where multimodal media content is used as stimuli for emotion elicitation to users, estimating the temporal causality from the various stimulus modalities is a difficult task. For example, considering a movie clip as a stimulus, it may contain varying sources of information, such as actor facial expressions, actor speech, background music, the scenery that is depicted, the text that is articulated by the actors, etc. Until now, emotion recognition approaches may receive multimodal inputs and infer affective states, but the problem of determining which specific cues of the multimedia content caused that specific affective state, is far from solved[300].

Societal and media industry drivers

Vignette 1: Using affective analysis to avoid sensationalism in news coverage

Anna is an editor in a big national newspaper that covers the civil war in another country. While her aim is to ensure the public is aware of the issues, she wants to ensure that the material is objectively portraying the situation without use of sensationalist material that would be against the ethical code of conduct. As a large number of articles come through every day, a tool that helps understand where each piece stands in terms of emotional neutrality would help her more accurately curate the material. Furthermore, if such a tool could also localise the emotionally heavy text extracts and suggest less charged language it would greatly help editors in the paper to maintain the standard set by the ethical code. Similarly, with video reports of events automatic analysis of content could help either broadcasters or independent regulators ensure sensationalism and emotionally charged news do not air at the expense of accuracy.

Vignette 2: Producing media content that captures audience engagement and attention

[297] J. Yang, X. Gao, L. Li L, X. Wang, J. Ding, "SOLVER: Scene-Object Interrelated Visual Emotion Reasoning Network", IEEE Transactions on Image Processing. 2021 Oct 19;30:8686-701

[298] D. She, J. Yang, M. M. Cheng, Y. K. Lai, P. L. Rosin, L. Wang, "WSCNet: Weakly supervised coupled networks for visual sentiment classification and detection", IEEE Transactions on Multimedia, 2019 Sep 5;22(5):1358-71

[299] Z. Xu, S. Wang "Emotional Attention Detection and Correlation Exploration for Image Emotion Distribution Learning", IEEE Transactions on Affective Computing, 2021.

[300] T. Mittal, P. Mathur, A. Bera, D. Manocha "Affect2MM: Affective analysis of multimedia content using emotion causality", In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021 (pp. 5661-5671)

Kate is a director working for a media production company that focuses on educational content for children. Given their intended audience, it is essential for the content they produce to be highly engaging in order to maintain children's attention. This is a very difficult challenge, given that the content must also be informative and leverage attention to achieve educational objectives. Furthermore, it is important that the emotions the content stimulates in the audience must not be harmful or traumatising, which is a concern when subject matter is sensitive in nature.

This problem can be addressed with focus groups consisting of the audience in question (in this case children), or by judgment calls by adults. The latter method is, naturally, less dependable, while the former can be expensive, time consuming, and imprecise, depending on the size of the focus group and how accurately it represents the target audience. AI tools measuring the engagement and affective response could facilitate this process by providing estimates of the impact of each piece of content on the audience, and, ideally, identify ways to retain attention and invoke engagement that are also productive and educationally beneficial.

Future trends for the media sector

Below, we highlight some of the emerging opportunities for research in the field of affective analysis to be translated into applications in the media sector:

- Creating tools that can identify and flag content that might trigger intense or undesirable emotional responses from the audience. Such tools could be used for **content moderation** across a variety of platforms and content modes (audio, visual and text), and could help shield the audience from problematic content that might otherwise evade other filters.
- Integrating affective analysis tools in **interactive media** (e.g. video games) would open the possibility for content producers to adjust their creations to the responses of the audience in real time and in a personalised manner. For example, video games might adjust their gameplay and/or visual and audio based on a player's emotional feedback to make them feel more challenged, be engaged, or to increase their immersion in the game's narrative.
- Developing more fine-grained and complex **multimedia tagging systems**, taking into consideration recent research findings about the multi-dimensional distribution of emotional states evoked by multimedia stimuli.[301,302]
- Measuring the audience engagement and tracking behavioural changes during the presentation of multimedia content (e.g. movies), can be used to train ***models with predictive capabilities***, so as to estimate future ratings of a movie[303].

[301] A. S. Cowen, D. Keltner, "Self-report captures 27 distinct categories of emotion bridged by continuous gradients", Proceedings of the National Academy of Sciences, 2017 Sep 19;114(38):E7900-9

[302] J. J. Sun, T. Liu, A. S. Cowen, F. Schroff, H. Adam, G. Prasad, "EEV: A Large-Scale Dataset for Studying Evoked Expressions from Video", arXiv preprint arXiv:2001.05488

[303] R. Navarathna, P. Carr, P. Lucey and I. Matthews, "Estimating audience engagement to predict movie ratings", IEEE Transactions on Affective Computing, 2017, 10(1), pp.48-59.

- Building **recommendation systems** for multimedia content, based on user preferences that are estimated from neurophysiological signals[304].
- Generation of novel media content (alternatively to the selection of pre-existing content) to be streamed based on user preferences, combining neurofeedback techniques and generative machine learning approaches[305].

Goals for next 10 or 20 years

The long term goal of Affective Computing in analysis and understanding of affect requires progress in several fronts, including in the theory of emotions and in understanding of the role they play in the way that we perceive the world and ourselves, in the way that we interact with others and in the way we act in the world. Affect manifests in widely varied ways, depending on the context - this varies from prototypical facial expressions, to subtle changes in the behaviour and neurophysiological responses. In this sense a holistic, multimodal approach that models the interplay between different human signals is required. To this end progress in ***sensing and data collection equipment at large scale*** *is* essential. In this direction, technological developments that will provide equipment capable of capturing modalities such as MEG and fMRI in out-of-lab environments[306, 307] would be needed in the next 5 years. Beyond that, a major goal in this direction for the next 5-10 years would be data collection at large scale, and the development of models that learn from this data collectively, so as to build personalised models in a way that respects privacy.

Recognition of emotions in others requires a theory of mind; that is an understanding of people's goals and their perception of the world. In this sense, progress in the field of Affective Computing requires progress in ***Machine Perception***, including among others in the field of computer vision, natural language processing and audio and speech analysis so as to model the world and the people in it. A goal with a 10-20 years horizon would be the development of artificial intelligence systems that would be able to ***model people's goals and intentions*** and be able to assist them in achieving them.

6.5 NLP and conversational agents

Contributors*: Adrian Popescu (CEA), Julien Tourille (CEA)*

[304] S. Koelstra, et al., "DEAP: A Database for Emotion Analysis; Using Physiological Signals". IEEE Trans. Affect. Comput. 3(1): 18-31 (2012)
[305] M Spapé et al., "Brain-computer interface for generating personally attractive images", IEEE Transactions on Affective Computing, 2021, PP. 1-1.
[306] T. Horikawa, A. S. Cowen,D. Keltner, Y. Kamitani "The neural representation of visually evoked emotion is high-dimensional, categorical, and distributed across transmodal brain regions", iScience, 2020 May 22;23(5):101060
[307] M. Liu, N. van Rijsbergen, O. Garrod, R. Ince, R. Jack, P. Schyns, "Semantic Decoding of Affective Face Signals in the Brain is Temporally Distinct", Journal of Vision, 2021, Sep 27;21(9):2589

Current status

Natural Language Processing (NLP) research has benefited from the introduction of deep learning models in the early 2010s. New methods and algorithms have unlocked possibilities and increased dramatically the performance on every task in the field over the last decade (see Figure 70). Specifically, the transfer learning paradigm has become the norm in the field. Large language models are trained on raw textual content scraped from the web and are fine-tuned on target tasks (e.g. named entity recognition or relation extraction). Among other model architectures that have been developed, the ***transformer architecture***[308] is the most versatile.

This new paradigm has allowed performance increases in various tasks such as named ***entity recognition and linking*** (the task of recognising entity mentions - e.g. the city of Paris - and assigning them a unique identifier - e.g. a Wikipedia URL), ***opinion mining*** (extract someone's opinion on one or several aspects of a topic) and ***argument mining*** (automatic identification and extraction of the structure of inference and reasoning of an argument).

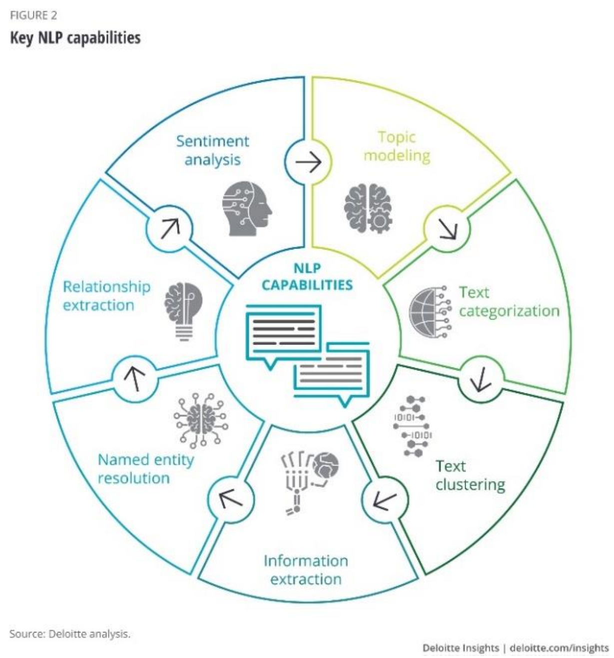


Figure 70: NLP capabilities. Image by Deloitte Insights.[309]

Research on **chatbots**, also called *conversational agents*, has also benefited from this new situation. Chabots have become ubiquitous in modern societies. They are applied in various domains such as health, customer service, education and office work. They also have become

[308] Vaswani et al. (2017). Attention is all you need. In: NIPS
[309] Image from Deloitte Insights: W. D. Eggers, N. Malik, and M. Gracie, Using AI to unleash the power of unstructured government data: https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/natural-language-processing-examples-in-government-data.html (Image source: https://www2.deloitte.com/content/dam/insights/us/articles/4815_AI-unstructured-data/figures/4815_Figure2.png)

part of our daily lives with the introduction of voice assistants on our connected devices (e.g. Apple's Siri, Amazon's Alexa or Microsoft's Cortana) (Figure 71).

Although academic research on chatbots has been active for the past 50 years[310], research and application on this topic have witnessed an increase in interest with the development of deep learning methods in the last decade. These new methods have improved the performance of algorithms for user intent and sentiment extraction, which are key elements in chatbot design. We have also observed the emergence of new models for natural language generation, trained on large conversational datasets. They generate accurate answers in natural language to user queries.

Figure 71: AI is listening.[311]

Research challenges

Although there was a huge leap forward over the last decade in NLP, several challenges remain. There are growing concerns around the use of large language models. They tend to ***reproduce societal biases*** that are present in the training corpus and have a ***large carbon footprint*** due to their sizes. More specifically, models such as GPT-3 are trained with huge volumes of Internet data, which are usually produced in the richest countries, in languages with higher linguistic footprint, and by communities with large representation. This results in models that fail to capture the culture of minorities and underrepresented groups and which may eventually discriminate against such groups[312]. In addition, their increased computational needs have a considerable environmental impact (e.g. it is estimated that training OpenAI's GPT-3 model produced the equivalent of 552 metric tons of CO_2, which is the equivalent of driving to the moon and back). Another interesting avenue of research is the integration of knowledge into

[310] Weizenbaum et al. (1967). *Contextual understanding by computers*. In: Communication of the ACM

[311] Illustration from XKCD at https://xkcd.com/1807/

[312] K. Hao, MIT Technology Review, We read the paper that forced Timnit Gebru out of Google. Here's what it says (2020): https://www.technologyreview.com/2020/12/04/1013294/google-ai-ethics-research-paper-forced-out-timnit-gebru/

deep learning models. There are indeed a lot of resources that are manually curated (e.g. UMLS or Wikipedia) and that contain valuable information.

Few-shot learning, i.e. the possibility to train a machine learning model with only a few examples, is also an active area of research (see section 5.3). Being able to train new models rapidly with low annotation costs could improve the diffusion of those models outside the academic world. To understand the size and resources needed to train some of the largest language models, we note that OpenAI's GPT-3, with its 175 billion parameters, was trained with 499 billion tokens and would require $4,600,000 to train - even when utilising the lowest priced cloud GPUs on the market[313].

Concerning the research on conversational agents, Følstad et al.[314] have identified several challenges. While performance of methods and algorithms has improved over the last couple of years, extracting precisely the ***user intent and sentiment*** remain difficult and conversational agents may have breakdowns when dealing with everyday situations. Then, adaptation to the user ***conversational context*** has been identified as one of the principal challenges in the early years and has remained an active area of research. Conversational agents should adapt to the social context and literacy competences of the users. Finally, standardised evaluation of systems and algorithms need to be implemented for chabots. Open and distributed datasets for evaluation have become a key element for sound and reproducible research in NLP for many years.

Societal and media industry drivers

Vignette 1: Multi-lingual document analysis and summarisation for investigative journalism

Adèle is a journalist who writes about the economic situation in developing countries, with particular focus on energy matters. To ensure a good coverage of this highly dynamic topic, she should have swift access to large amounts of news, social media contributions and corporate documents written in different languages. She needs a set of tools which: (1) translates domain-specific documents from potentially under-resourced languages to German, English and Croatian, the reporting languages by Adèle, in a reliable manner; (2) clusters them according to both topics and opinions about these topics to give an overview of the different points of view expressed; (3) mines arguments used in the source documents and evaluates their plausibility; and (4) summarises these documents. Such NLP-based tools would greatly facilitate Adèle's documentation work and would allow her to focus on the creative side of her reporting.

Vignette 2: Conversational agent for interactive and personalised audio content delivery in broadcast

Hector is in charge of the media delivery team of a public broadcaster. One of the innovative services of the broadcaster is to offer interactive and personalised audio content to their

[313] C. Li, OpenAI's GPT-3 Language Model: A Technical Overview (2020): https://lambdalabs.com/blog/demystifying-gpt-3/
[314] Følstad et al. (2021). *Future directions for chatbot research: an interdisciplinary research agenda*. In: Computing.

audience. Hector's team develops a conversational agent which adapts the content served to the user depending on their historical preferences but also on the go, based on questions asked when the content is played. The agent exploits cutting-edge NLP techniques in order to: (1) formalise historical preferences and propose material which mixes new content on topics that were already explored by the users and new topics which are likely to interest them; (2) understand the user's queries and integrate them in the dialogue model in order to be able to modify the content at any moment depending on the interaction with the user; (3) combine content from the broadcaster's archives and from open sources in a coherent material which is served to the user.

Future trends for the media sector

NLP has witnessed a true revolution during the last few years with large language models like GPT-3. NLP is expected to become increasingly mainstream in the media business though applications such as conversational agents and virtual characters, creative writing, robot journalism, interactive storytelling, voice search for image/video/audio, sentiment analysis in social media, voice dubbing, or multi-lingual translation.

On the news content production side, NLP is already in the newsroom, with large news broadcasters having their own research and development divisions. Their principal objective is to devise tools to smooth the editorial process. Among those tools, ***faceted search engines*** help journalists to look for information in databases, archives and on the web. They allow for rapid information gathering to help staff to write and edit news.

There is no extensive use of state-of-the-art NLP yet. But this may change in the future. Several academic research teams are devising ***automatic fact-checking systems*** that are able to tell whether the information is legitimate or spurious. ***Automatic argument mining*** is also actively pursued. Being able to extract the argumentation structure of a source document could help journalists to skim rapidly several sources of information.

Concerning information extraction, new methods for ***event, entity and relation extraction*** could help journalists to rapidly extract relevant documents from large databases. For instance, NLP techniques were already used to skim the large corpus available in the Pandora papers.

On the other side of the spectrum, news consumers could also benefit from the development of NLP. Automatically characterising news topics and opinions could help citizens to broaden their views and counter the effects of news recommendation systems that tend to only suggest content aligned with user's opinions.

All other media industry sectors are also expected to be impacted by the deployment of NLP algorithms. Video games, especially those involving open worlds, may benefit from conversational agents powered by AI to improve the gaming experience and offer a unique story to each participant. This also opens interesting opportunities for the Metaverse where bots could potentially interact with users' avatars.

The film and TV industry can use new NLP techniques to automatically analyze audience reviews for fast opinion mining surveys or to predict audience engagement and movie success based on script analysis. Assisted multilingual subtitling or voice dubbing are also interesting applications for this sector while automatic or assisted translation in different languages can be a game-changer for the publishing industry.

Goals for next 10 or 20 years

NLP is a large and old academic research domain and has experienced several breakthrough changes over the last two decades. It is therefore difficult to draw objectives for the far future. However, we identified several interesting research avenues for the next few years.

As mentioned in previous sections, a lot of work has been done in representation learning, with the emergence of the transfer learning paradigm. Models have gained in complexity and ingenuity, and helped to improve the state-of-the-art for all tasks in NLP. There is an actual trend in devising more and more complex systems with billions of parameters. However, little research has been done on data centric approaches where the objective is ***to improve the overall representativity*** of the dataset that is fed to the models to increase downhill performance.

Furthermore, although new methods and datasets have been devised over the past couple of years, probing the knowledge that is embedded in these models remains difficult and needs further research. A better understanding of the content of these models could lead to improvement in other NLP areas.

Finally, the emerging topic of ***model hybridation***, i.e. combining existing curated knowledge and unsupervised language model, needs further investigation. Although recent language models manage to capture knowledge in an unsupervised fashion, they fail to learn fine-grained relations within the text.

6.6 Content moderation

__Contributors__: Georgi Kostadinov (Imagga), Chris Georgiev (Imagga), Pavel Andreev (Imagga), Ralitsa Golemanova (Imagga)

During the last few years, ***content moderation*** powered by Artificial Intelligence has grown exponentially. It has developed in ways that were unimaginable just a decade ago, breaking concepts and widening the horizons of what’s possible.

Automated content moderation has been fuelled by ever-evolving machine-learning algorithms that constantly improve in accuracy and speed. Just 10 years ago, image recognition was only able to classify and detect basic objects and shapes. Now, thanks to the advancements of deep learning, image recognition algorithms for instant detection of all types of inappropriate visual content are a reality.

Automated (also referred to as semi-automated) content moderation thus offers important new capabilities for businesses of different venues that need effective screening of digital content. The AI moderation platforms address a number of key challenges that online platforms and companies face, including:

- Huge amounts of user-generated content need to go online immediately, but still have to be monitored for appropriateness, safety, and legality. This can make it difficult for online platforms to grow and scale internationally if they don't have an effective way to ***screen all postings*** — textual, visual, and even live streaming. Without moderation, these businesses risk great reputational harm, along with other negative consequences.
- Content moderation has to happen ***in real-time***, which is especially difficult for live streaming and video that are becoming the most popular content formats. The complexity of screening visuals, texts and moving images at the same time is tremendous.
- ***User safety*** and especially the ***protection of vulnerable groups*** is becoming a priority in legislation that covers digital platforms. This means that in many places across the globe, online businesses are required by law to have solid Trust and Safety programs and protection mechanisms based on content moderation. This is necessary not only to ensure the upholding of their internal principles and guidelines, but to safeguard consumers.
- The ***stress and harmful effects on human moderators*** from exposure to shocking, violent and disturbing content is significant. Digital businesses aim to minimise these negative consequences and to protect their moderating teams from the worst content.
- Digital platforms have to be able not only to scale in terms of countries and amounts of content that goes live, but also to adapt to ***quickly changing circumstances and norms for content appropriateness***.
- Public ***manipulation, political propaganda, disinformation*** through fake news, and the rise of deepfakes are disturbing yet prevailing new phenomena online. Both official authorities and online platforms need an effective way to fight them, and machine learning algorithms are the key to that.

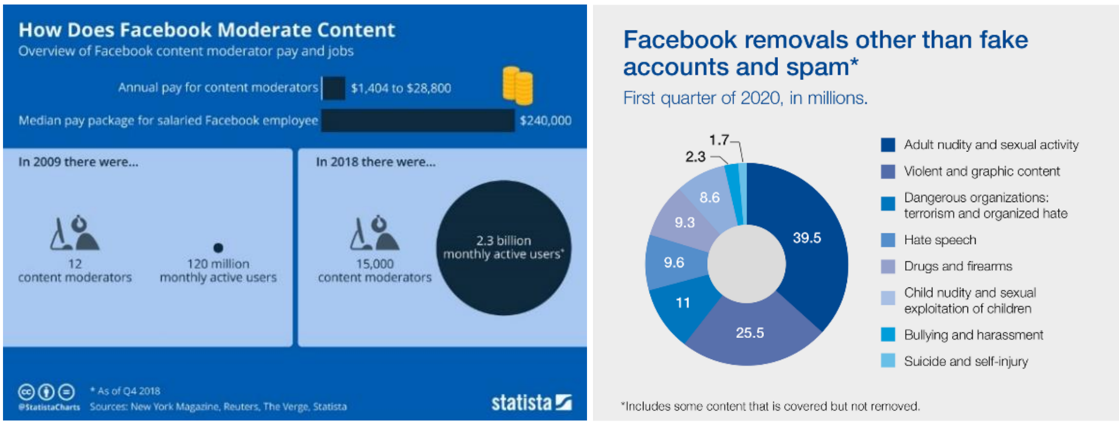


Figure 72: Some interesting moderation statistics: a) Facebook content moderation between 2009 and 2018[315], and b) Facebook content removals per category in 2020[316].

Research challenges

While holding great potential and already showing impressive results, there are challenges that AI-powered content moderation is facing.

One of the major issues with which automated content moderation is struggling ***is recognising context***. Machine learning algorithms can find it difficult to differentiate between subtle cultural and social trends and phenomena. For example, if the algorithm is set to remove all nudity, this is what it would do - even if the nudity is related to art or important news pieces. A prominent example was the case from 2016 when Facebook removed the photo of the iconic Vietnamese 'napalm girl' who is naked[317].

A growing body of ***national and international regulations*** are already affecting online platforms that allow the sharing of user-generated content and communication between users. The regulations are steadily pushing for effective content moderation in order to protect people and ensure a fair and safe environment for all.

The European Union is gradually moving forward with its Digital Services Act[318], and so are individual countries like the United Kingdom, France, Germany, and the United States with their national legislation. These new regulations include provisions to provide adequate protection of users against inappropriate and harmful content. The rationale of these legal requirements

[315] Image source: Statista, How Does Facebook Moderate Content (2019) https://www.statista.com/chart/17302/facebook-content-moderator/

[316] Image source: P. Barrett, Who Moderates the Social Media Giants? (2020): https://static1.squarespace.com/static/5b6df958f8370af3217d4178/t/5ed9854bf618c710cb55be98/1591313740497/NYU+Content+Moderation+Report_June+8+2020.pdf (p. 10)

[317] S. Levin, J. Carrie Wong and L. Harding, "Facebook backs down from 'napalm girl' censorship and reinstates photo", The Guardian (2016): https://www.theguardian.com/technology/2016/sep/09/facebook-reinstates-napalm-girl-photo

[318] European Commission, Proposal for a Regulation on the European Parliament and of the Council on a Single Market for Digital Services (DSA Act) and amending Directive 2000/31/EC, 15 December 2020, COM(2020) 825 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN

seems understandable, especially considering the infamous cases of crimes and murders that have been live streamed on major platforms like Facebook, YouTube, and Periscope[319].

On the other end, this might create more fears of ***moderation bias and censorship***, resulting in limiting the freedom of expression, and the right of the user to know why their content was removed. What is clear is that we need to better understand the context, the political situations in different regions of the world as well as cultural and religious particularities and properly transfer that knowledge to any AI algorithm that will deal with automated content moderation.

Another important challenge that AI platforms need to overcome is ***multilingual and multicultural moderation***. While they are getting better at it and are surely improving how content moderation in different languages is performed, there are still obstacles in the way. The process is not only about acknowledging the direct meaning of words and phrases, but their ***social and cultural connotations*** that may make them offensive and inappropriate. In this respect, the more feedback machine learning algorithms receive, the better they can become at spotting the nuances in content — which is definitely not a mission impossible, but simply a gradual process that takes processing large amounts of data.

Live streaming and live video are another interesting challenge for AI-based content moderation platforms. They generate such a substantial amount of data per second that manual moderation is simply an impossible task. Moreover, applying AI on each frame of the live broadcast generates high platform costs. A fast and accurate AI needs to be developed to overcome these hurdles and reach efficient and cost-effective moderation of live streaming and videos.

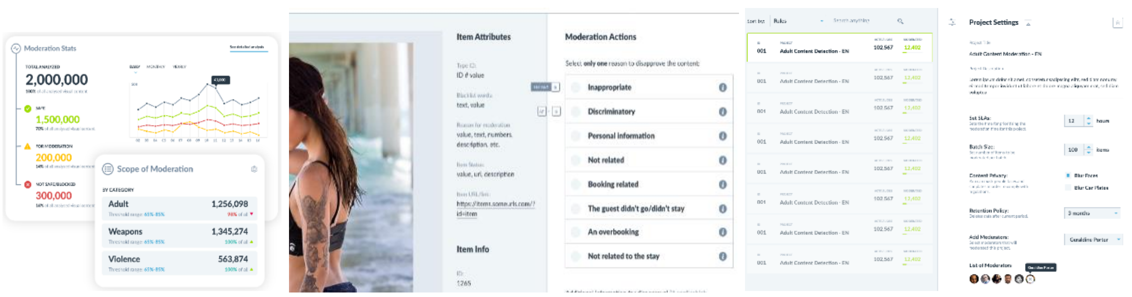


Figure 73: A content moderation tool by Imagga[320].

Societal and media industry drivers

Vignette 1: Automatic moderation of content harmful for vulnerable communities

Merry is a manager in an online newspaper with a solid and long-standing reputation for monitoring content that might be harmful for vulnerable communities. Her job is to ensure every piece of information that gets published through the media's channels meets company

[319] O. Solon, "Why a rising number of criminals are using Facebook Live to film their acts", The Guardian (2017): https://www.theguardian.com/technology/2017/jan/27/rising-numbers-of-criminals-are-using-facebook-to-document-their-crimes

[320] Image source: https://imagga.com/content-moderation-platform (provided by Imagga)

standards for content that might be shown to vulnerable groups. She needs to check not only facts and textual references, but also visuals. In a way, Merry is a modern-day fighter against discrimination – one of the plagues of today's digital world.

But doing all of this on her own – even with her fact-checking and moderation teams – is a monstrous task. The process simply takes enormous amounts of time and effort. That's why Merry needs a viable and scalable solution for checking and preventing the spread of inappropriate visuals.

This is especially important in some parts of Europe and other countries of the world, where protection of basic human rights is not fully enforced and problematic content can easily slip in. Ordinary citizens, and sometimes politicians share problematic content for their own gain – and it's very difficult to sift through what's acceptable and what's not. Merry has to figure out how true the lead is and whether to publish it.

With the help of an AI-powered content moderation platform, Merry can screen various materials around the topic for authenticity. She can catch textual references, as well as photos that have already been posted, for example.

Vignette 2: Real-time moderation of live-streaming content

Daniel is a content editing manager at an online news outlet. He's in charge of guaranteeing that all published content complies with the standards of the media and the legal framework. His most challenging task is to ensure the compliance of live streams. Catching inaccurate and harmful video content in real time is a tough nut to crack. Without technical support, it is a burdensome task to monitor live streaming content as it occurs.

This is where a content moderation platform based on machine learning algorithms can kick in. It processes all live video streams, checking them for inappropriate verbal and visual elements. If there are such, the platform can immediately signal this to the editors.

Daniel can test the capabilities of the AI platform in practice during an important live streaming with a local politician at a public rally. The situation is uncontrollable, as it's a place full of people where anyone can appear and take over the stage. With the help of the moderation platform, Daniel can have the live stream screened for problematic content throughout the whole event.

Future trends for the media sector

The role of content moderation in the media sector cannot be overestimated — in fact, it's crucial for its wholesome development on a couple of levels:

- Content moderation algorithms are the key to fighting the *online disinformation*. They can spot inaccuracies in textual data and fake or old visuals and videos. These capabilities can be a game changer for the media sector that direly needs adequate fact-checking in the ocean of information that gets published daily.

- Protecting vulnerable groups, under-represented communities and women as well as fighting terrorism will be of utter importance in the future and media companies need to find ways to better understand such content and adopt or develop tools to address it.
- On the basis of the massive volumes of content that machine learning algorithms process, they can also make *predictions about the types of content that needs to be moderated*. This can be of huge importance for combating harmful tendencies in user-generated content sharing.
- Content moderation, if done in-depth, can help in detecting the intent of disinformation in order to differentiate between positive (for example, to keep state secrets) and negative (to influence opinions and harm society).
- With the fast adoption of Metaverse, a new set of moderation requirements are emerging as this new medium merges reality with virtual worlds. Fake and authentic will have a totally new meaning and content moderation platforms of tomorrow need to adapt to this new trend of virtualisation of almost everything.

Goals for next 10 or 20 years

The long-term vision for AI-powered content moderation is a truly ambitious one.

First and foremost, content moderation would need to leave the semi-automatic status it currently has. To be fully useful, scalable, and powerful, it should be ***more autonomous***. The vision is that AI-powered content moderation platforms would become a monitor that is always on and oversees any and all content that goes online. It would screen for all types of abnormalities to ensure protection of users from harmful and illegal content, and a safe online environment without fake news — thus taking care of everything from violence and nudity to propaganda, radicalisation and disinformation, and all that's in between.

The second big goal for content moderation's evolution is ***self-learning*** — which is already in motion, but can reach new heights. With the data that is being fed in the moderation platform in real time, the machine learning algorithm becomes better and better. It expands its knowledge base with practical examples and input from moderators. With time, this is how the AI can become more independent from humans in terms of feedback loop. In the foreseeable future, this can reach a point where the moderation platform becomes an autonomous machine that identifies and filters content accurately and effectively with no human input.

A third long-term goal for content moderation is the creation of ***instant and efficient on-device moderation***. Nowadays, moderation is executed on the server end, only after a piece of content has already gone live. This means that harmful content can sneak in for a moment and be accidentally shown to users. In the near future, moderation would be possible on the user device itself. This would happen before the content has gone live. This advancement would enable the prevention of illegal and disturbing content appearing on the device level, thus ensuring full protection for end users.

7 Trustworthy AI: future trends for robust, interpretable, privacy-preserving and fair AI

The analysis of section 3 showed that there are many security and societal risks involved with the adoption of AI by the media industry (and by many other domains). This can essentially be summarised as to whether AI can be trusted to behave ethically, lawfully and offer accurate results. Concretely, this means that AI models should be transparent about their decision-making processes, should not discriminate, should respect our privacy, and should not be vulnerable to attacks.

Trustworthy AI aims at providing a framework for the development of ML technologies, which guarantees their suitability with respect to the democratic and ethical values shared of our society (Figure 74). This emerging field of AI is typically divided within four broad dimensions, namely ***AI robustness***, ***AI explainability***, ***AI fairness***, and ***AI privacy***. In section 7, we focus on these four aspects of Trustworthy AI, attempting to examine the current status, applications, challenges and future trends of Trustworthy AI for the media industry.

A separate subsection is dedicated to each dimension, following the same format that was adopted for section 5 (see the introductory paragraph of that section).

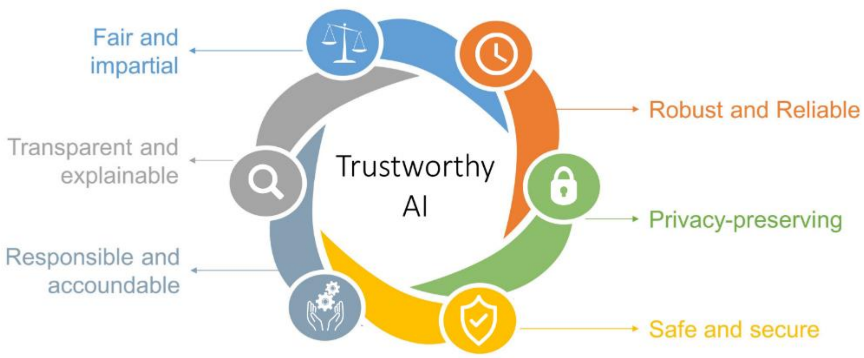


Figure 74: Aspects of Trustworthy AI[321].

7.1 AI robustness

***Contributors**: Killian Levacher (IBM)*

Current status

AI has become ubiquitous in almost every industry. The media sector specifically sees journalists increasingly relying on AI-enhanced tools for information retrieval, fact checking, or image verification to avoid deepfakes, to name just a few examples. Besides the many advantages of

[321] Figure adapted/adapted from Deloitte's Trustworthy AI™ framework (source: https://www2.deloitte.com/us/en/pages/deloitte-analytics/solutions/ethics-of-ai-framework.html)

AI, it also introduces novel vulnerabilities and ways for applications to fail, including unexpected model behaviour because of sensitivity to naturally occurring distributional shifts in the input data, or specifically crafted adversarial inputs aiming to influence or control the AI method. Researchers are identifying and categorising an increasing number of adversarial threats, techniques, and tactics against AI[322], which malicious attackers exploit to subvert the AI. A **lack of robustness in AI** can have consequences from **reputational damage** to **serious physical-world injuries**.

Specifically crafted adversarial perturbations and interactions against the security and privacy of AI, often described as attacks, can broadly be categorised into evasion, poisoning, extraction, and inference. Figure 75 shows how these attacks relate to a simplified AI training pipeline including the ML model and its training dataset and the following sentences will limit itself to cite one of many representative papers on each topic.

Evasion attacks[323] attempt to craft adversarial inputs, often imperceptible to humans, which fool AI into making wrong decisions. For example, deepfake videos can be modified with adversarial perturbations to mislead AI-based deepfake detectors.

Poisoning[324] attempts to modify, or better poison, the training dataset used to train AI models, for example, to introduce backdoors in the trained model that can later be used to control the AI or to degrade the success of entire training pipelines.

Extraction[325], or model-theft, aims to learn the model parameters and architecture to rebuild similar or sometimes even exact copies at a fraction of the cost of the original, often proprietary, model. Next to the monetary loss, the extraction of models can contribute to other security threats, for example if the stolen copy of a model is used to craft stronger adversarial examples. This is of high concern for businesses which have invested in creating datasets, employ experts and maintain computational resources to build state-of-the-art AI models.

Finally, **inference** attacks[326] attempt to find leaks of information that is contained in the training data by only accessing the trained AI model. Such leaks are particularly problematic where sensitive personal data, e.g., health records, have been part of the training data and individuals could be identified.

---

[322] MITRE ATLAS - Adversarial Threat Landscape for Artificial-Intelligence Systems: https://atlas.mitre.org/

[323] Croce, F., and M. Hein. "Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks." *International conference on machine learning*. PMLR, 2020.

[324] Aghakhani, H., et al. "Bullseye polytope: A scalable clean-label poisoning attack with improved transferability." *2021 IEEE European Symposium on Security and Privacy (EuroS&P)*. IEEE, 2021

[325] Jagielski, M., et al. "High accuracy and high fidelity extraction of neural networks." *29th USENIX Security Symposium (USENIX Security 20)*. 2020

[326] Choquette-Choo, C.A., et al. "Label-only membership inference attacks." *International Conference on Machine Learning*. PMLR, 2021.

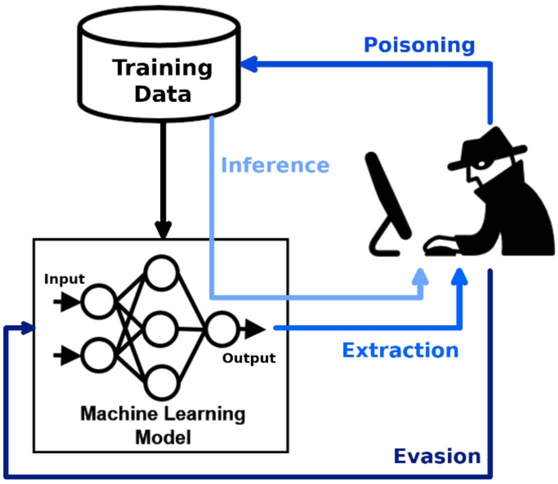


Figure 75: Different types of attacks against machine learning models.[327]

Research challenges

In response to AI attacks, the field of ***Robust AI*** has now gained a lot of attention. This field aims at identifying defense mechanisms by which AI systems can be made more robust to such attacks. This field is of critical importance if we are to continue increasing our reliance on machine learning systems. Within the field of Trustworthy AI, it could even be considered the foundation of all other sub-disciplines since the value offered by Explainable AI or Fair AI systems vanishes if the public cannot be certain these have not been tampered with by attackers.

Current research challenges in this field include the development of novel approaches to defend AI and establish adaptive, scalable robustness evaluation methods of AI and defensive methods. Defending AI is not trivial and the number of attacks is growing rapidly[322]. Here, we highlight a few promising approaches and cite representative articles tackling current challenges.

One of the most successful approaches against evasion is ***adversarial training***, which uses adversarial examples of the training samples during training to increase the bounds of robustness[328] and continuous research is focusing on improving its efficiency and strength[329]. ***Defences at various steps of the AI pipeline*** including pre-processing inputs[330] and post-processing outputs[331] are promising but correct evaluation of their true robustness in white-box

---

[327] Image source: Adversarial Robustness Toolbox - https://adversarial-robustness-toolbox.readthedocs.io/en/latest/_images/adversarial_threats_attacker.png

[328] A. Madry et al., “Towards Deep Learning Models Resistant to Adversarial Attacks”. *arXiv preprint arXiv:1706.06083*, 2017

[329] E. Wong“Fast is better than free: Revisiting adversarial training”. *arXiv preprint arXiv:2001.03994* (2020)

[330] P. Samangouei, “Defense-GAN: Protecting Classifiers Against Adversarial Attacks Using Generative Models” by P. in *arXiv preprint arXiv:1805.06605* 2018

[331] T. Lee “Defending Against Machine Learning Model Stealing Attacks Using Deceptive Perturbations” *arXiv preprint arXiv:1806.00054* (2018)

scenarios is not trivial[332] and requires critical thinking, careful analysis of the defence and the development and application of adaptive attacks[333]. ***Defences in black-box scenarios***, where attackers target models only with query access, seem more successful[334,335] and are focusing on reducing the number of model queries allowed to attack. ***Defences for generative models***, which until recently had received relatively less attention by the adversarial community, are also being developed[336]. ***Certification of robustness*** with mathematical guarantees have been presented[337] and challenges include the applicability to different types of perturbations[338] and increasing the certified robust perturbations.

Societal and media industry drivers

Vignette: AI robustness for investigative journalism

Jane is a journalist working for a prestigious news agency which, as member of an investigative news consortium, reports frequently on international diplomacy around the world. Her agency is known for investigating and reporting corruption at the highest levels of government which resulted in diplomatic scandals affecting the stability of nation-states.

Jane's professional ambition is to become the author of a news article describing such a scandal. As it turns out, she recently got contacted by an anonymous source who shared with her details of a secret meeting which just occurred last month between two authoritarian world leaders. As proof this meeting did occur, the anonymous source attached a set of photos showing both leaders around a table and shaking hands.

This information confirms international suspicions that the two governments have been secretly preparing a nuclear attack. This could very well become the scoop of the decade Jane has been waiting for her entire career.

She decides to go ahead and write an article about it. Before doing so however, she uses her agencies suite of AI fact checking tools to double check that the material sent to her is indeed genuine. Among others, her agency possesses state-of-the-art re-identification AI and deepfake detection tools, which can respectively authenticate the identity of an individual on a photo and guarantee that it is not deepfake material. All the AI fact checking tools return positive results on all checks. The photo indeed depicts both leaders at a meeting and the material is authentic.

---

332 N. Carlini et al. "On Evaluating Adversarial Robustness". *arXiv preprint arXiv:1902.06705* (2019)
333 F. Tramèr et al., "On Adaptive Attacks to Adversarial Example Defenses", *Advances in Neural Information Processing Systems* 33 (2020)
334 H. Li et al. "Blacklight: Defending Black-Box Adversarial Attacks on Deep Neural Networks". *arXiv preprint arXiv:2006.14042* (2020)
335 S. Chen, "Stateful Detection of Black-Box Adversarial Attacks", in *Proc. of the 1st ACM Workshop on Security and Privacy on Artificial Intelligence*. 2020
336 A. Rawat A, K. Levacher, M. Sinn, "The Devil is in the GAN: Defending Deep Generative Models Against Backdoor Attacks" in *arXiv preprint arXiv:2108.01644* (2021) and presented at Blackhat USA 2021 https://www.youtube.com/watch?v=zBZbBMvuPXU&t=153s
337 J. Cohen, E. Rosenfeld, and Z. Kolter. "Certified adversarial robustness via randomized smoothing." International Conference on Machine Learning. PMLR, 2019
338 H. Salman, "Certified Patch Robustness via Smoothed Vision Transformers". *arXiv preprint arXiv:2110.07719* , 2021

She goes ahead and writes an article describing the meeting and the serious international diplomatic implications resulting in this event. Before publishing the article, her editor decides to inform and share her draft with the consortium's sister agencies.

The next day her editor opens her inbox and discovers that one of the sister agencies, using the exact same set of AI fact checking tools has detected that the photos about to be published are Deepfakes. The editor re-runs an analysis of the photos on her side which still confirms the photos are authentic and decides to contact the AI team in her agency for help.

The next day the AI team informs Jane and her editor that the AI fact checking tool suite of nearly all news agencies part of the consortium have been the target of a major cyber-attack by a famous group of secret hackers.

These hackers have tampered the AI models used by nearly all agencies part of the consortium for months without anyone noticing, which explains why different results were returned. As it turns out, the AI teams in these agencies had not updated their AI models suite with the latest adversarial defences which left them vulnerable to such attacks. The editor immediately requests for all article publications to be halted temporarily until all model defences are up to date.

Jane is at first disappointed that her source turned out not to be genuine. However, after a few minutes of reflection, she then realises that a secret organisation hacking the most prestigious investigative news consortium, is an even bigger story to be told. It turns out she did have the scoop of her life after all, just not the one she had initially thought.

Future trends for the media sector

The recently proposed EU Legal Framework for AI (AI Act)[339] explicitly mentions evasion and poisoning attacks and proposes that the creators and deployers of AI systems should be responsible to be aware of the AI's limitations and potential harms, deploy state-of-the-art mitigating measures, and be able to explain and reproduce the AI's actions.

The type of damages that adversarial attacks could cause in the media industry is endless. The list below presents a subset of concrete scenarios, which will become increasingly relevant over the next decade.

Social networks and recommendation engines, have already received a lot of attention in recent months, based on the filter bubble effects they produce which can affect real world events such as elections. It has now become clear that such properties can be deliberately leveraged by hostile foreign agents[340] with devastating consequences by simply creating fake accounts and *indirectly* manipulating the recommendations. Hence, the ability to *directly* target

[339] European Commission, Artificial Intelligence Act: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206
[340] BBC News, "Russia 'meddled in all big social media' around US election" (2018): https://www.bbc.com/news/technology-46590890

the recommendation models themselves, without even creating a single fake account, using adversarial attacks could produce an even greater and more precise damage.

Another area in which various media industries will be increasingly vulnerable to adversarial **poisoning attacks** will be those relying upon generative models. The quality of synthetically generated **text and audio** has dramatically improved over recent years. Such models are already being used in **the movie**[341] **and game industry** for voice over or dialogue generation for instance. As a result, this opens the door to poisoning attacks in which generative models could be subverted in producing harmful content such as hate speech or any other material, which could directly harm consumers or at a minimum significantly tarnish the reputation of a company

Investigative journalism within the near future will increasingly require the processing of very large datasets of various types (e.g. Panama papers[342], Lux leaks[343]) in order to discover potential abuses of power. Since these datasets are very large, the use of machine learning models to infer and highlight salient insights will be inevitable. As a result, organisations using such models will be vulnerable to **inference attacks**. Private innocent individuals whose data happened to be contained within such datasets could be exposed by such attacks and lead to defamation legal cases against the news organisation.

The multimillion **movie industry**, which has always been at the forefront of innovation, will increasingly require to protect itself against adversarial attacks. So as to reduce production costs, mechanisms which enable the automation of parts of the production process are constantly been created. Models that automatically generate movie trailers are now being used for that purpose[344]. As investments in such models increase, so will the quality of generated trailers. This will represent major savings in production costs and thus give a big advantage to production companies with quality models. As a result, the temptation to steal such models through **extraction attacks** will only increase as the industry adopts such models in their pipelines.

Finally, the **advertisement industry** isn't immune to adversarial attacks either. Most of the publishing industry relies on funds derived from AI models carefully selecting and placing individual advertisement in the most relevant pages of a website. Such models could be vulnerable to poisoning attacks, which could purposely place advertisements in undesirable locations (e.g. a petroleum advertisement placed right next to an article describing an oil spill from that company) leading to significant loss in revenue for both the publisher and advertiser as well as reputational damages.

---

[341] H. Rosner, "The ethics of a deepfake Anthony Bourdain voice", The New Yorker (2021): https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice
[342] W. Fitzgibbon, "The Panama Papers: Exposing the Rogue Offshore Finance Industry", International Consortium of Investigative Journalists (2021): https://www.icij.org/investigations/panama-papers/
[343] The Irish Times, "Lux Leaks": https://www.irishtimes.com/business/lux-leaks
[344] I. Fadelli, "A new model that automatically generates movie trailers", TechXplore (2021): https://techxplore.com/news/2021-11-automatically-movie-trailers.html

All of the examples presented above, increasingly lead to the necessity for AI models to provide some form of **certification** to the public **regarding their trustworthiness**. Efforts such as the AI Factsheets initiative[345] are already underway. These AI model certifications provide the means for various consumers to verify whether a given model has been thoroughly tested as well as identifying the potential risks involved in using it. For a given model, these certifications provide this information in various forms, each adapted to the knowledge of relevant stakeholders (data scientist, journalists, product manager etc.). Among the many attributes these certificates can contain (a large set of example certificates for diverse models can be browsed here[346]), attributes could consist of the list of attacks which a given AI model was tested against along with the relative drop in accuracy, the list of defences used for the protection of the model against such attacks, the type of bias for which a model will be most vulnerable to etc. As the industry becomes increasingly reliant on AI models, we can expect the list of such attributes to grow over time as well as the legal requirements to provide such certificates to consumers.

Goals for next 10 or 20 years

The development of robust AI will always be a competition between attackers and defenders in which the AI community must constantly stay ahead of malicious attackers by discovering and reporting vulnerabilities and developing better defences[322], similarly to traditional IT security where new vulnerabilities, viruses and malware are constantly discovered and defences in form of patches and updated anti-virus databases are providing protection.

We can expect that in the next 5-10 years novel approaches for secure AI will reach a level of maturity that covers most of the common datasets and applications. They will be easy to apply by common data scientists and to integrate within existing pipelines. Moreover, for AI in general to continue increasing its reach in society, it is imperative that the development speed gap, currently witnessed between that of attacks and defences, is reduced. Adaptive defence mechanisms, which can dynamically adjust their defence algorithms based on evolving and adapting attacks will be developed.

7.2 Interpretable AI

***Contributors**: Mara Graziani (HES-SO), Henning Muller (HES-SO)*

Current status

The research field in ***AI interpretability*** has grown very quickly in the last decade. The literature on interpretability techniques counts (until 2020) more than 70,000 papers containing either

345 M. Hind, "IBM FactSheets Further Advances Trust in AI", IBM Research Blog (2020): https://www.ibm.com/blogs/research/2020/07/aifactsheets/

346 IBM AI FactSheets 360: https://aifs360.mybluemix.net/

"XAI", "explainability", or "interpretability"[347]. As Doshi-Velez and Kim argue[348], the rising interest in "***opening the black-box***" is motivated by the fact that evaluating the performance of complex machine learning models is an ill-posed problem, and that the sole model accuracy on the test is not sufficient to describe the model's inner functioning and the satisfaction of important desiderata. In applications where making a mistake would have strong implications on people's lives (e.g. credit allowance, insurance premiums, healthcare, etc.), the work on interpretable (or explainable) AI has emerged as a way to provide individuals with insights about automated decision-making. The main concept of explainable AI is illustrated in Figure 76.

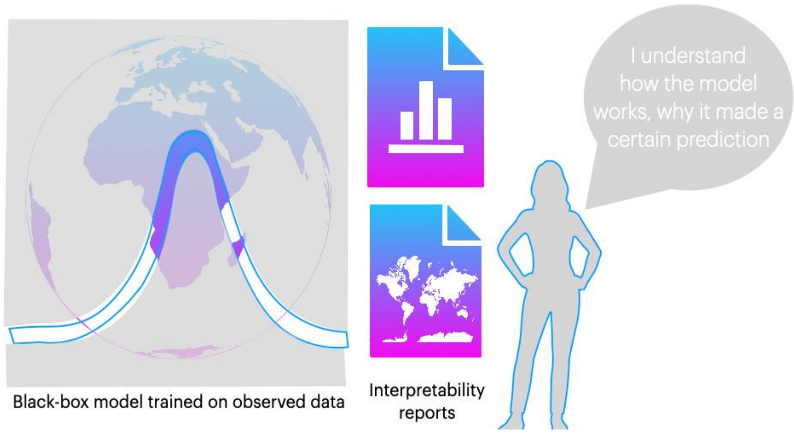


Figure 76: Explainable AI – the concept.

The EU's General Data Protection Regulation (GDPR)[349], in effect since May 2017, has officialised the need for ***reliability in addition to accuracy*** of the models, where reliability includes the generations of explanations that assign meaning to AI decision-making, improving the user's mental model of the automated process. Interpretability thus stands as a part of the social interaction between the AI system and its user. As we would expect bankers to explain why they rejected a loan, doctors to explain why they decided to discontinue treatment, and politicians to explain why they wanted to implement a certain policy, we would expect AI systems to justify their decision making if it impacts our lives. The GDPR forms, in this context, a "***right to be informed***", by claiming: (i) the right not to be subject to automated decision-making and safeguards enacted thereof (Article 22 and Recital 71); (ii) notification duties of data controllers (Articles 13–14 and Recitals 60–62); and (iii) the right to access (Article 15 and Recital 63). The information provided about the AI system must be meaningful to the individual confronted with the automated decision.

Some of the technical terms used to distinguish most of the current approaches to obtaining interpretable AI were introduced by Lipton[350]. Two important distinctions commonly adopted in

[347] From app.dimensions.ai, as accessed in August 2021.
[348] F. Doshi-Velez and B. Kim. "Towards a rigorous science of interpretable machine learning." arXiv preprint arXiv:1702.08608 (2017).
[349] General Data Protection Regulation (GDPR): https://eur-lex.europa.eu/eli/reg/2016/679/oj
[350] Z. Lipton, "The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery." Queue 16.3 (2018): 31-57.

the domain are that of (i) local vs. global explanations and (ii) built-in vs. post-hoc methods. ***Local explanations*** refer to explanations that are only true for a single input. ***Global explanations****,* on the contrary, explain the model behaviour for an entire set of inputs, e.g. all images of a single class in the dataset. ***Built-in methods*** introduce interpretability as one of the objectives of the model optimisation function. These methods are included in the more general notion of intelligible AI. An example is that of inherently interpretable models, e.g. linear regression, where the linear increase of a feature value corresponds to a proportional increase in the model output. ***Post-hoc methods*** are methods that generate explanations without requiring the retraining of the model parameters with interpretability constraints.

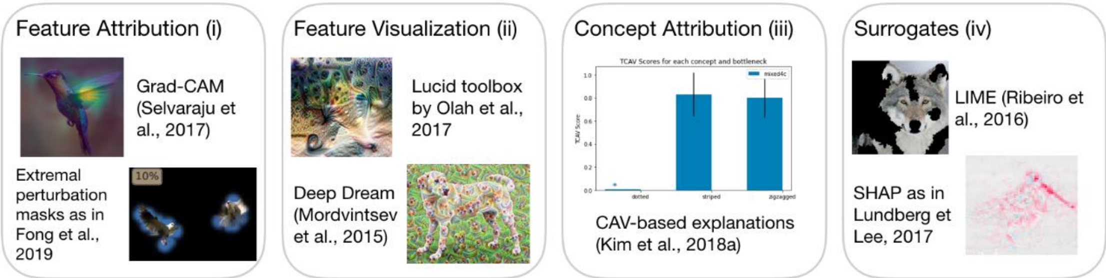


Figure 77: Post-hoc Interpretability approaches for AI models.[351]

Post-hoc approaches can be further grouped (as shown in Figure 77) depending on the form of the generated explanations into: (i) ***feature attribution methods***, that aim at identifying the input features that are the most relevant to the prediction; (ii) ***feature visualisation***, that aims at uncovering the patterns that are learned by intermediate layers and units; (iii) ***concept attribution***, that explains the model outcome in terms of high-level concepts and (iv) ***surrogate explanations***, namely those techniques that use a proxy model (generally simpler to interpret) to generate explanations. Two additional strategies are case-based and textual explanations. Another important distinction is between model-agnostic and model-dependent models. ***Model-agnostic methods*** do not need any access to the internal model's logic and/or state (e.g. model parameters), and only rely on the input and output pairs. They consider the model to be interpreted as a black box where only the output for a given input is observable. As a result, model-agnostic methods can be applied to all models. Perturbation methods such as occlusion and Local Interpretable Model-agnostic Explanations (LIME) are model-agnostic.

Most of these techniques still require large testing with users. Most often, human beings often do not need complete causal chains of explanation and may prefer a trustworthy account of understandable reasons expressed in clear and simple language. Collaborations should thus be built to develop types of human-computer interactions in ML that are more understandable to non-ML experts.

Research challenges

[351] Image taken from M. Graziani. Interpretability of Deep Learning for Medical Image Classification: Improved Understandability and Generalization. Diss. University of Geneva, 2021

The main research challenges related to AI explainability can be summarised as follows:

- There is ***no ground truth of what constitutes a good explanation*** on real-world problems. While for controlled datasets it may be possible to identify the set of variables that are relevant to the data generation process, when dealing with the complexity of the data describing real-world phenomena such as the spreading of tweets and news, it is impossible to know a-priori the set of sufficient features to describe the phenomenon. Explanations, besides, may be differently understood by people with different backgrounds. Because of the subjectivity of the receivers of explanations, it is hard to define what an optimal explanation should be.
- The ***ability of understanding is highly subjective***. Computer scientists may understand information differently from experts in other domains. The analysis of the impact of subjectivity on the explanations is a major challenge. Anthropology should help understanding what characters, e.g. journalists, fact-checkers, content-creators, would define an "understandable" explanation.
- ***Interactive explainability*** techniques should be provided to allow humans to work together with AI systems. The explanations should then be related to the practices and the context. If a journalist were to use a fact-checking tool using AI, for example, to determine the reliability of a piece of information, then the fact-checker reliability should also be evaluated through explanations that are related to the practice and context of the journalism.
- Explainability should relate to ***eventual causation links*** between the features learned by the model and the features that are actually used to make the prediction. A convolutional neural network may be learning, for example, to extract visual features of texture, colour and shape from images depending on the architecture being used and how the filters, skip connections, and invariances are encoded in the model. Despite learning such features, only some of these may then be used to make the prediction. The aim of causal analysis is thus to evaluate not only *what features are being learned*, but also *how these features are used by the model* and *how changing such features would impact the classification decision*.

Societal and media industry drivers

Vignette: An interactive and explainable fact-checking tool for journalists

Glenn has been commissioned to verify the reliability of the viral spreading of information about the "*beginning of human cloning in some European countries as a means to promote the availability of organs for transplants*". He is strongly convinced that this information is false, being against the ethical rights and policies currently established by the Union. However, he can find confirmation of this news on multiple reliable sources. Puzzled, he decides to use a multimodal AI system with a content-retrieval module to verify the credibility of this information by cross-matching all the existing evidence. The AI system retrieves and backtracks all the available information online about the news. Several content from yet un-regulated social media is filtered out and automatically reported as a non-cross-validated information on true-or-fake.socialmedia.com. As expected, the AI system clarifies that the information is a fake.

Relieved, Glenn starts digging about the AI outcome to uncover the reasons for this prediction. The system provides an interacting interface where it is possible to explore the information that was used to make the prediction in multiple ways. Some input images retrieved from online are highlighted in some regions pointing to the time of the information publication. Some checkered artifacts are pointed in several videos. Surprisingly, the model highlights as very important apparently unrelated documents in multiple languages that report power shortages at multiple media sites, including BBC, CNN and several European national televisions. At this point, Glenn interacts with the model to further illustrate the connection between the power shortages and the prediction of the news as fake. The AI system presents Glenn a cascading analysis of the events that influenced the prediction. The documents on the power shortage all document the same time as the actual electricity interruption. No major security issue was reported, so this information had been overlooked. The videos and articles about the beginning of human cloning, however, all reported a publication time within the power shortage. The clear signs of corruption in the images such as the checkered artifacts show that the videos were actually generated by other AI methods. Because of all of these reasons, the model concluded that it is fake news. Glenn now has more relevant questions to verify, namely whether there was a synchronised attack to major media industries to finally diffuse fake information again.

This was the first case of viral spreading of fake information again in the past 15 years, since the creation of true-or-fake.socialmedia.com and the use of explainable AI to detect misinformation.

Future trends for the media sector

Some of the opportunities concern the use of interpretable AI as a means to empower the possibilities of journalists, fact-checkers, film-makers, content providers, content creators, advertisers and other figures in the media. Interpretability may be used to evaluate on the basis of what reasons some content was predicted as fake. Disparities in the automated decision making due to bias towards gender, race and income will be highlighted by the explainability tools to promote equity and inclusivity. Interpretability may be used to identify bias and discrimination in models trained on incomplete and unbalanced datasets. For example, it may be used to guarantee that the tools for automatic casting in the film industry do not penalise non-white people. Similarly, it may be used to evaluate the presence of negative bias towards the recommendation of LGBTQ contents.

Goals for next 10 or 20 years

The future will see AI applications as part of a partnership with humans for empowering their capabilities. Knowledge must be collected on how humans acquire new information and represent their beliefs. Interpretable AI shall consider this knowledge to provide decision support. The systems shall be interactive, so that they may adapt to the user's needs and clarify any unclear explanations with further elaboration. Systems may directly point out societal risks such as biases, discrimination and misinformation as soon as possible. The AI user of the future is informed at all times and builds trust in the machine by interacting with the system.

7.3 Privacy-preserving AI

Contributors*: Patrick Aichroth (FhG-IDMT), Thomas Köllmer (FhG-IDMT)*

Current status

"If you give me six lines written by the hand of the most honest of men, I will find something in them which will hang him."— Cardinal Richelieu[352]

Privacy can be understood as the ability of individuals to control their personal information, the right to not be observed, to be left alone, and to keep relationships and personal matters secret. It helps us making our own decisions, without being observed or disturbed, keeping social pressure at bay.

AI has created new challenges for privacy: Machine learning requires large datasets for training, creating fantastic new possibilities, but also pushing an increased desire for collecting data, including personal data to be used for targeted advertisement or service improvements. We create and share large amounts of personal data e.g. by using smartphones but also passively while living in an environment that collects more and more data – think about CCTV, online payments, location tracking. The problem is, thanks to the technologies being developed and applied within recently, we are neither fully aware of the kind and amount of data being collected and used, nor can we predict how data will be used by AI in the future.

Does that mean that privacy is not a concern for people anymore? Probably not. A survey in the EU indicated that 41% of respondents were not willing to share personal data with private companies, only 5% were willing to share facial images or fingerprints with private companies, and 55% were afraid of criminals or fraudsters accessing their personal data.[353] Similarly, a 2022 survey in the US indicated that a majority of respondents was concerned about how much data is collected about them by companies and the government (79%/64%), believing that much of what they do online and on their cellphone is being tracked by companies or the government (72%/47%). More than 80% said that they feel having very little or no control over the data collected about them and that the potential risks outweigh the benefits when it comes to companies collecting data.[354] Such concerns could also explain some of the worries regarding AI, for instance in Germany, where public skepticism about AI is already considered a key burden to innovation by German SMEs.[355]

---

352 Fischer, D. H. (2009). Champlain's Dream (Reprint Edition). Simon & Schuster, p. 704

353 How concerned are Europeans about their personal data online? (2020, June 15). European Union Agency for Fundamental Rights. https://fra.europa.eu/en/news/2020/how-concerned-are-europeans-about-their-personal-data-online

354 Auxier, B., & Rainie, L. (o. J.). Key takeaways on Americans' views about privacy, surveillance and data-sharing. Pew Research Center. Abgerufen 4. February 2022: https://www.pewresearch.org/fact-tank/2019/11/15/key-takeaways-on-americans-views-about-privacy-surveillance-and-data-sharing/

355 Rammer, C. (2021). Herausforderungen beim Einsatz von Künstlicher Intelligenz, Ergebnisse einer Befragung von jungen und mittelständischen Unternehmen in Deutschland. Mannheim: Bundesministerium für Wirtschaft und Energie (BMWi)

However, the view that privacy and AI are mutually incompatible is both wrong and dangerous: if we are willing to give up privacy because we (wrongly) believe we have to, we are paying a tremendous price. And if we bluntly reject technologies such as AI because we (wrongly) believe we have to give up privacy, we give up on designing and improving technologies, ending up with a self-fulfilling prophecy, leaving markets to those who do not care about privacy. Instead, we should aim at solutions to build privacy into AI, and invent and use technologies that allow us to do that. But which technologies could that be?

As for protecting privacy for data analysis, ***anonymisation techniques*** and ***concepts*** have traditionally played a key role, including ***k-anonymity***, ***l-diversity***, and ***t-closeness***. Within a dataset, there are *direct identifiers* (attributes which directly identify an individual), *quasi-identifiers* (attributes which can identify an individual if combined with other quasi-identifiers, although the definition is not always used consistently[356]) and *sensitive attributes* (attributes that shouldn't be linkable to an individual, e. g. information about religion, politics, health, etc.). ***K-anonymity*** is about ensuring that there are at least k entries with the same attribute combination (e. g., 2-anonymity ensures there are at least two entries), by removing or altering data, e. g. by applying *suppression* (replacing values with standard values), or *generalisation* (replacing individual values with broader categories or ranges). ***L-diversity*** and ***t-closeness*** are extensions of *k-anonymity*, addressing some of its weaknesses, such as homogeneity and background knowledge attacks.[357] All of them aim at the goal of addressing ***re-identification risks,*** i.e. "*the potential that some supposedly anonymous or pseudonymous data sets could be de-anonymized to recover the identities of users*."[358], and the domain is often referred to as ***Privacy-Preserving Data Publishing***.

Privacy-Preserving Data Publishing remains useful in many domains, but with the advent of AI, it does not seem sufficient anymore: The mentioned approaches and statistical techniques are designed to consider a limited number of selected attributes, but AI is about processing large amounts of data with high dimensionality and complexity, resulting in several new challenges: much higher likelihood of sensitive information being included, much higher likelihood of models being able to reveal sensitive information, and significantly increased difficulty in protecting sensitive information.

Research challenges

Privacy-Preserving AI (PPAI) is about addressing the specific challenges related to AI and privacy, which can be split into four categories (see also section 7.1):[359]

---

[356] Bettini, C., Wang, X. S., & Jajodia, S. (2006). The Role of Quasi-identifiers in k-Anonymity Revisited. arXiv:cs/0611035. http://arxiv.org/abs/cs/0611035

[357] Machanavajjhala, A., Gehrke, J., Kifer, D., & Venkitasubramaniam, M. (2006). L-diversity: Privacy beyond k-anonymity. 22nd International Conference on Data Engineering (ICDE'06), 24–24.

[358] Chia, P. H., Desfontaines, D., Perera, I. M., Simmons-Marengo, D., Li, C., Day, W.-Y., Wang, Q., & Guevara, M. (2019). KHyperLogLog: Estimating Reidentifiability and Joinability of Large Data at Scale. Proceedings of the 2019 IEEE Symposium on Security and Privacy.

[359] Thaine, P. (2020). Perfectly Privacy-Preserving AI. Medium. https://towardsdatascience.com/perfectly-privacy-preserving-ai-c14698f322f5

- ***Training Data Privacy***, which is about preventing malicious actors from reverse-engineering the training data.
- ***Input Privacy***, which is about preventing that a user's input data can be observed by other parties, including the model creator.
- ***Output Privacy***, which is about preventing that the output of a model is visible to anyone except the user whose data is being inferred upon.
- ***Model Privacy***, which is about preventing that the model is stolen.

Some of the most relevant attacks in the context of privacy and security include ***inference attacks***, i.e. attacks that aim at analyzing data to gain knowledge about a subject, and ***model poisoning***, i.e. attacks that manipulate data in order to influence or corrupt the model.

Among inference attacks, ***input inference attacks*** (also referred to as *model inversion* or *data extraction*) are probably the most common and relevant from a privacy perspective. Such attacks aim at extracting data from the training dataset, e.g. obtaining attributes, or audio or image training data related to a person based on her name. Similarly, ***membership inference*** and ***attribute inference*** attacks aim at finding out whether a particular example was in the dataset. It is noteworthy that the latter can not only be used as an attack, but also to check whether privacy-preserving measures were applied for training.

One of the most important approaches within the realm of privacy-preserving AI is ***Differential Privacy*** (DP). First introduced in 2006[360], it provides a mathematical definition of privacy that ensures that no individual data entry (e.g. referring to a specific user) has significant influence on the overall output distribution and hence no significant influence on query results. This is typically achieved by adding noise to the input, the output, or by modifying the query algorithm itself (Figure 78). With DP, the amount of information that can be gained about a given individual is limited to a specific value, thereby also providing a way to measure privacy. At the same time, overall accuracy does not significantly decrease (the statistical properties of a dataset are preserved). However, the costs and benefits of DP depend on the specific case at hand: the smaller the datasets, the more accuracy tends to decrease due to the added noise, while for larger datasets, accuracy may even *increase*, as the introduction of noise can reduce overfitting. Due to its advantages, especially regarding application in the context of AI and cloud computing, DP has seen a significant increase in relevance and demand within recent years, with DP libraries being developed and used by many major companies and vendors.[361] Moreover, further adaptations and variants to DP have been developed e. g. for specific needs related to AI

[360] Dwork, C. (2006). Differential Privacy. In M. Bugliesi, B. Preneel, V. Sassone, & I. Wegener (Hrsg.), Automata, Languages and Programming (S. 1–12). Springer. https://doi.org/10.1007/11787006_1

[361] Examples of Differential Privacy libraries: https://github.com/IBM/differential-privacy-library (IBM), https://github.com/OpenMined/PipelineDP (Google / OpenMined), https://github.com/opendp/smartnoise-core (Microsoft), https://github.com/pytorch/opacus (Facebook)

training, including *Differentially Private Stochastic Gradient Descent* (DPSGD)[362,363] or *Private Aggregation of Teacher Ensembles* (PATE)[364].

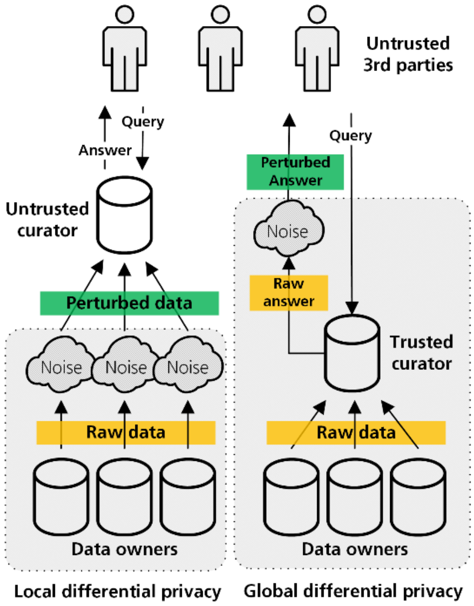


Figure 78: Local and global differential privacy.

Homomorphic Encryption (HE) represents another key technology in the PPAI domain: such encryption schemes can perform different classes of computations over encrypted data and can be split into *partially* and *somewhat* homomorphic encryption (which are limited with respect to operation type / amount), and *fully homomorphic encryption* (which can perform addition and multiplication any number of times). In the context of PPAI, HE is a powerful tool e.g. in that it can support data processing and training performed by an aggregator, without the aggregator gaining access to the clear-text data (Figure 79). One of the most relevant HE approaches is *the Cheon-Kim-Kim-Song* (CKKS) scheme[365], which has been implemented within several libraries[366] and is subject to standardisation activities.[367] One key challenge for the practical implementation of HE is computational and memory overhead, which varies significantly depending on the scheme and implementation used, but challenges also include practical security challenges such as key management. *Secure Multiparty Computation* (SMPC) provides yet another relevant

[362] Song, S., Chaudhuri, K., & Sarwate, A. D. (2013). Stochastic gradient descent with differentially private updates. 2013 IEEE Global Conference on Signal and Information Processing, 245–248.
[363] Wu, X., Li, F., Kumar, A., Chaudhuri, K., Jha, S., & Naughton, J. F. (2017). Bolt-on Differential Privacy for Scalable Stochastic Gradient Descent-based Analytics. arXiv:1606.04722 [cs, stat]. http://arxiv.org/abs/1606.04722
[364] Papernot, N., Song, S., Mironov, I., Raghunathan, A., Talwar, K., & Erlingsson, Ú. (2018). Scalable Private Learning with PATE. arXiv:1802.08908 [cs, stat]. http://arxiv.org/abs/1802.08908
[365] Cheon, J. H., Kim, A., Kim, M., & Song, Y. (2017). Homomorphic Encryption for Arithmetic of Approximate Numbers. In T. Takagi & T. Peyrin (Hrsg.), Advances in Cryptology – ASIACRYPT 2017 (S. 409–437). Springer International Publishing.
[366] e. g. HElib (IBM): https://github.com/homenc/HElib, SEAL (Microsoft): https://github.com/microsoft/SEAL
[367] Homomorphic Encryption Standardization. https://homomorphicencryption.org/standard/

cryptographic technique for PPAI. It can be used to jointly compute a function over inputs while keeping inputs private, serving as an addition to the aforementioned techniques.

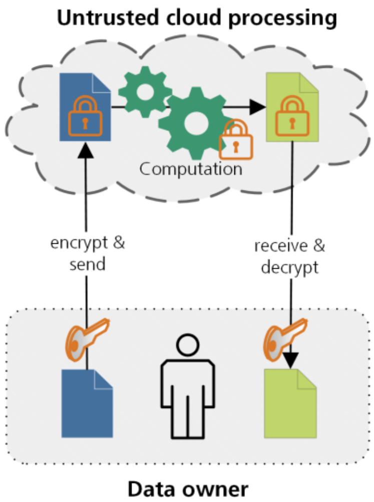


Figure 79: Homomorphic Encryption.

Another key tool for PPAI tools is ***Federated Learning*** (FL). First introduced in 2017[368], FL aims at conducting the training process among several participants, but without the need to exchange the training data: the data can remain "on prem", which means that FL provides great potential for many applications with respect to elevated security, copyright and privacy requirements (Figure 80). There are different variations to FL, ranging from centralised FL (with a central, orchestrating server) to decentralised FL (nodes / participants are able to organise themselves). As for the other PPAI techniques mentioned, practical application of FL is not trivial, and related challenges depend on the specific application context.[369] Also, for many applications, it is necessary to complement FL with HE and DP, as FL alone cannot prevent attacks such as inference attacks (to be addressed with HE and DP) or model poisoning (to be addressed with DP).

Finally, especially in the context of media, PPAI requires tools that can be used to remove person-related information from audio-visual material, e.g. by using source separation[370] or speech alienation[371] in the case of audio material.

[368] McMahan, B., Moore, E., Ramage, D., Hampson, S., & Arcas, B. A. y. (2017). Communication-Efficient Learning of Deep Networks from Decentralized Data. Proceedings of the 20th International Conference on Artificial Intelligence and Statistics, 1273–1282

[369] A Comprehensive Survey of Privacy-preserving Federated Learning: A Taxonomy, Review, and Future Directions: ACM Computing Surveys: Vol 54, No 6. (o. J.). Abgerufen 4. February 2022

[370] Hennequin, R., Khlif, A., Voituret, F., & Moussallam, M. (2020). Spleeter: A fast and efficient music source separation tool with pre-trained models. Journal of Open Source Software, 5(50), 2154.

[371] Liang, D., Song, W., & Thomaz, E. (2020). Characterizing the Effect of Audio Degradation on Privacy Perception and Inference Performance in Audio-Based Human Activity Recognition. In 22nd International Conference on Human-Computer Interaction with Mobile Devices and Services (S. 1–10). Association for Computing Machinery.

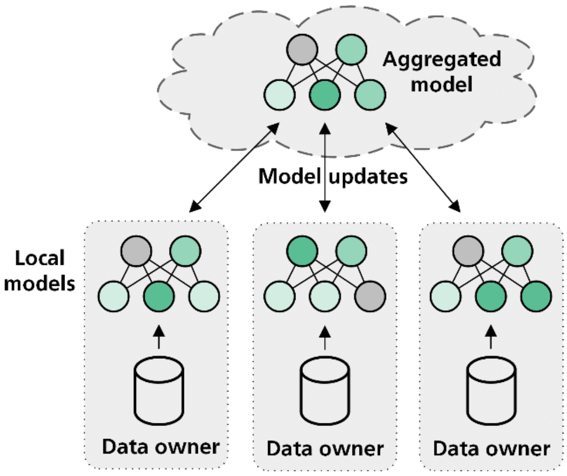


Figure 80: Federated Learning.

Societal and media industry drivers

Vignette 1: Privacy-enhanced news recommendations

Linda is very interested in politics and news. She is constantly looking for new tools that can help her receive news about the topics relevant to her, while covering a broad range of political perspectives: she is aware of the dangers of echo chambers and likes to read “other” opinions and news sources from time to time, even (or especially) if she does not agree with them, also to reflect her own arguments – she is active in a political party and therefore often involved in discussions. Finally, she finds a cross-vendor recommendation service for this, which uses her feedback to provide personalised news recommendations. It is a great service, providing high-quality information. However, Linda fears that the service learns a lot about her political views and preferences. She is afraid what could happen if such information is leaked or published, and checks information and reviews about the service. She learns that the service applies a range of privacy-preserving technologies including federated learning, homomorphic encryption, differential privacy and various standard security protocols, to ensure that no one except herself learns about her choices and preferences, and that no information can be connected to her real identity. She also learns that all these privacy and security claims have been tested and certified by well-known, independent security and privacy companies. She has finally found what she has been looking for.

Vignette 2: Privacy-enhanced speech transcription

Joseph is the owner of a small software company and has been using an app- and cloud-based speech transcription service from a big non-European vendor service since years, using it for a lot of purposes that involve highly personal information. He has recently grown increasingly concerned about what data is stored by the service, after reading about potential attacks and recent hacks of other services. He knows that the service uses speech recordings to continuously improve its performance – he agreed to do that before installing it, as it significantly improved performance. But apart from that, he does not know much about the service, and did not understand the terms of use in detail but confirmed them anyway, because it was the only working app for this purpose and he urgently needed it. Now, he is completely unsure whether speech recordings from him were uploaded and used, whether and how they were stored, and whether and how the created transcripts are stored and protected.

Actually, he uses many cloud-based services, e.g. for collaborative editing, issue management, backup and other means, without really understanding all legal details and technical risks involved. He is not happy at all about that, because he has a feeling that this comes with significant business and personal risks, but he gave up on the topic because he found that even legal or technical expert friends were hardly able to do that assessment. He would be willing to pay for more trustworthy alternatives with comparable performance and usability, but for some kinds of applications, he is simply not aware that any such alternatives exist.

One day, while spending his holidays with his family, Joseph learns that there was a phone scam that tricked his colleagues to transfer and effectively lose 150,000 € based on a replayed voice recording or deepfake, similarly to what happened in other cases.[372] At the same time, he learns that he is blackmailed with sensitive business information to be published if he does not agree to transfer another 350,000 € within the next days. He is not sure about where the respective leak came from but rejects to transfer 350,000 €, resulting in a publication of information about his clients and his private life that results in a very tough period for him and his family, but also for his company, which almost goes bankrupt in the months after the incident.

After many months of introspection, Joseph decides to turn the terrible experience into an opportunity, resulting in a strategic shift within his company: it starts developing software for transparent, privacy-aware speech transcription and note management services. After two years of very intense research and development in a strategic collaboration with other organisations being specialised in AI-based transcription, recommendation and privacy enhancing technologies, Joseph's company is now offering such services with good usability and performance, easy-to-understand terms of use, and supporting multiple languages. The services are continuously audited with respect to their privacy/security promises by 3rd party companies, and turn out to be a great market success, especially among security-/privacy-aware companies.

Future trends for the media sector

Privacy will be key for many media applications, including the following domains:

Recommendation will play an increasingly important role in the media sector, considering the ever-increasing amount of data and the need to deliver relevant information to audiences. Considering an increasing awareness about privacy, users will demand more transparency about which of their data is used and how, requiring that state-of-the-art technologies are applied to

372 J. Damiani, "A Voice Deepfake Was Used To Scam A CEO Out Of $243,000", Forbes (2019): https://www.forbes.com/sites/jessedamiani/2019/09/03/a-voice-deepfake-was-used-to-scam-a-ceo-out-of-243000/

protect their information and that technical audits are performed to ensure that such technologies are used (and used properly).

The ***processing of audio-visual data*** will frequently require the need to remove person-related information before content is stored and processed, using appropriate technologies.

The ***fabrication of synthetic audio and video*** material will bring an entirely new challenge for privacy – after all, being confronted with fabricated information about oneself that others wrongly consider authentic is even worse than losing control over authentic personal information. New technologies aiming at the detection and localisation of synthetic audio-visual material are currently being developed, but more awareness and modified processes within the media industry to deal with this problem are also required.

Services and offers in the media domain, as in any other domain, will only be successful if they combine good performance / usability etc. and privacy requirements. In other words, privacy and other trust aspects can become key features for commercial success, but only if we combine regulation (which is great, but not sufficient) with innovation.

In order to develop new AI-based tools and services for the media sector especially in Europe, ***"clean" datasets*** will be needed. It is good that Europe emphasises the need for privacy protection, but what is also needed is support and investment in creating and providing the appropriate alternative datasets.

AI does not only pose challenges. It can also be used to ***support and protect privacy***. For instance, increased automation and the use of AI can reduce the risk of data loss due to human error, it can improve auditing of privacy weaknesses in systems, it can improve transparency about the use of personal information, and it can support humans in becoming much more aware and rational about making cost-benefit considerations about privacy, including assessment of long-term costs and benefits (which human bias makes us especially unfit for).

Goals for next 10 or 20 years

Anonymisation techniques will be used to remove person-related data from audio-visual material, and differential privacy will provide metrics and means to guaranteed levels of privacy. Secure multiparty computation, homomorphic encryption and federated learning will allow AI training and data processing without the risks of unintended data loss and privacy violations. Media companies and society will have developed the organisational, educational and technical means to deal with manipulated, fabricated and decontextualised information as a potential threat to privacy. Principles and regulation regarding privacy will always be intertwined with respective innovation and investments in technology, and privacy and AI will not be treated as mutually incompatible anymore, but instead, privacy will become an integral feature of AI-based products.

7.4 AI fairness

***Contributors**: Samuel C. Hoffman (IBM Yorktown Heights)*

Current status

Machine learning has the potential to bring about enormous change for good in the world. Already we see mature applications in language translation, medicine, image and speech recognition, and more. These technologies have the chance to democratise their benefits and reach incredible scales. But, as with any other powerful tool, machine learning can also serve to exacerbate inequalities and make historically powerful groups more powerful. This is the crux of the field of ***algorithmic fairness***: how to ensure the fruits of AI are shared equitably and discrimination is prevented.

Besides the obvious ethical implications, there are many reasons for machine learning application designers to consider fairness. In many fields, there are legal protections for disadvantaged groups such as equal employment, housing, and banking. In other cases, user trust can be eroded if a company is shown to have discriminated, even after the problems are fixed. Conversely, transparency around fairness can be a competitive advantage.

In recent years, many advancements in AI have been shown to demonstrate ***biases***. Many facial recognition technologies have been shown to perform significantly worse on dark-skinned faces[373] (see Figure 81). Healthcare applications have been shown to result in poorer outcomes for minority patients[374]. AI tools for job recruiting could have discriminated against women applicants[375]. Criminal recidivism models led to harsher penalties for black defendants[376]. In all of these cases, discrimination was almost certainly not the explicit goal of the model's producer but rather, a side effect of inattention to the potential harms.

So how do these models come to espouse biases if they are not intentionally designed that way? Unlike early rule-based AI models, machine learning and especially deep learning study very large sets of data in order to discover patterns in the relationship between inputs and outputs. Therefore, if a certain class of people are not well represented in the data, either in total number or proportional outcomes, a machine learning model may generalise poorly to real users. Furthermore, if the data used to train a model contains historical instances of bias from the real world, these tendencies can easily transfer to the model's outputs. Human decision-making processes can be flawed in the same ways, however, even if the model happens to be no worse

[373] Buolamwini, J., & Gebru, T. (2018, January). Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on fairness, accountability and transparency* (pp. 77-91). PMLR.
[374] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453.
[375] Dastin, J. (2018, October 10). Amazon scraps secret AI recruiting tool that showed bias against women. Reuters. https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G
[376] Angwin, J., Larson, J., Mattu, S., & Kirchner, L. (2016, May 23). Machine bias. ProPublica. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing

than humans, the speed, scale, and often lack of accountability (the removal of humans from the process) of AI means the overall effect will be more harmful.

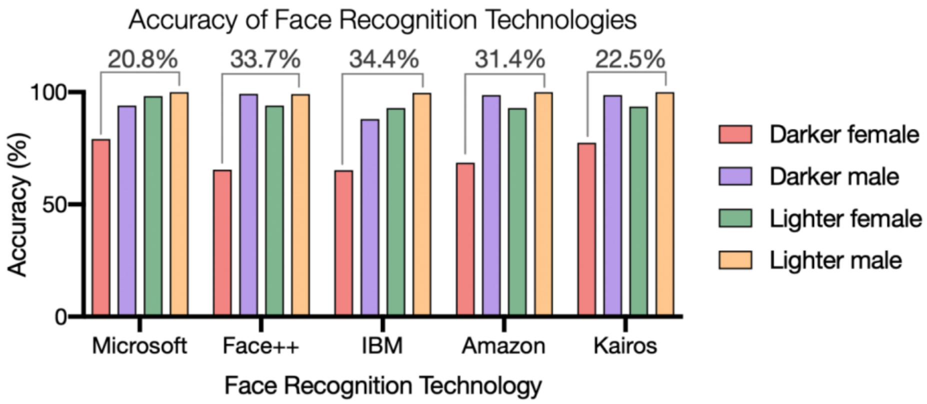


Figure 81: Results of the Gender Shades[377] project which audited various facial recognition models. This analysis prompted the targeted companies to improve their error rates on darker-skinned females by up to 30% within 7 months[378] (Image source: A. Najibi's blog post[379]).

But how can we tell if a machine learning model will have bias or compare the bias of different models? There is no one answer to this question since the relative value of tradeoffs varies based on the application. From an *individual* perspective, similar people should be treated similarly (measuring similarity between people is yet another tricky task, however)[380]. Likewise, changing uninformative aspects such as race or religion (a *counterfactual* example) should not result in different outcomes. From a *group* perspective, people from different sections of the population – grouped by a sensitive attribute such as gender, race, disability, belief, or age – should, on average, receive equal results. In some cases, the *procedure* or treatment of all individuals must be equivalent. These are all compounded by the fact that the real world often contains institutional and systematic disadvantages so in order to correct for these, measuring performance (e.g. accuracy) itself must take fairness into account.

It is easy to see how some of these definitions can be in conflict with each other. For instance, in a medical setting, following the same procedure for men and women would be inadvisable in many situations and yet we should still expect tests and surgeons to perform equally well on both. Less obvious, though, is the fact that even different group-based metrics are often

[377] Gender Shades: http://gendershades.org/overview.html
[378] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. In Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (pp. 429-435).
[379] A. Najibi, Racial Discrimination in Face Recognition Technology (2020): https://sitn.hms.harvard.edu/flash/2020/racial-discrimination-in-face-recognition-technology/
[380] Dwork, C., Hardt, M., Pitassi, T., Reingold, O., & Zemel, R. (2012, January). Fairness through awareness. In Proceedings of the 3rd innovations in theoretical computer science conference (pp. 214-226).

impossible to satisfy simultaneously[381,382]. Furthermore, Simpson's paradox[383] tells us we must carefully choose the groups of interest since bias can be hidden by grouping data differently. Thoughtful attention must be paid to which priorities are most important to various stakeholders in order to define fairness in a given situation.

Research challenges

Much research in the past few years has focused on mitigating biased outcomes in predictive AI models. These can generally be broken down into three categories: ***data mitigation or pre-processing*** which attempts to modify the training data so that future algorithms that use it will be less biased, ***fair estimators or in-processing*** methods which optimise for fairness and predictive performance concurrently, and ***prediction or post-processing*** which intervenes after a model makes a decision and changes it if bias is detected (Figure 82). Many proposed methods make theoretical guarantees about improving certain fairness metrics. It is important to remember, however, that improving the quality, size, and makeup of the original training data is usually preferred since mitigation generally comes with a tradeoff between accuracy and fairness.

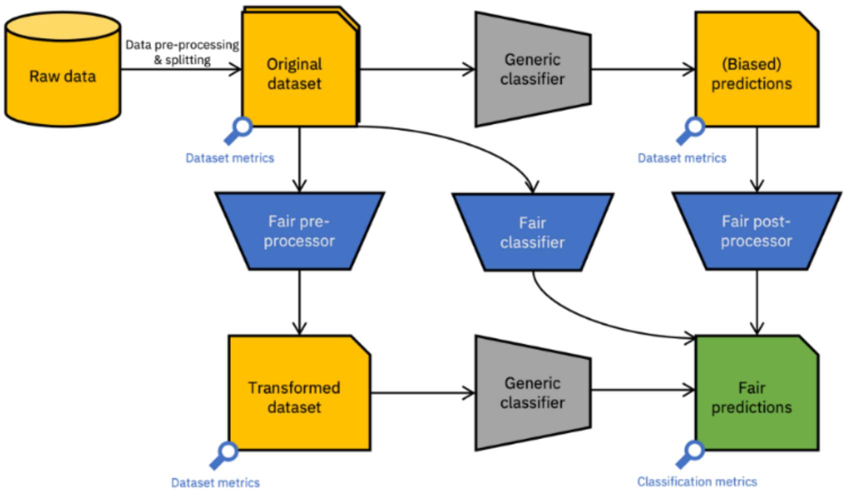


Figure 82: Various ways to mitigate bias in AI models. Bias should also be measured at multiple points in the AI lifecycle, including in deployment.

Data plays a key part in not only algorithmic fairness but machine learning at large. In fact, data is so valuable that transparency is often explicitly avoided as a business priority. Furthermore, ***data privacy*** continues to be of greater concern to the public (see section 7.3). However, the ***opaqueness of data*** also makes it more difficult to reproduce and assess AI models and requires trust in the institutions designing them. ***Data and model transparency*** is an area of focus for

381 Chouldechova, A. (2017). Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. Big data, 5(2), 153-163.
382 Kleinberg, J., Mullainathan, S., & Raghavan, M. (2016). Inherent trade-offs in the fair determination of risk scores. arXiv preprint arXiv:1609.05807.
383 Wikipedia, Simpson's paradox: https://en.wikipedia.org/wiki/Simpson%27s_paradox

researchers which aims to build trust in algorithms by divulging performance statistics (e.g. accuracy, fairness, robustness, etc.) and details about the development lifecycle. To this end, various forms of "factsheets" have been proposed for data and models to document this information in a consistent and comprehensive manner[384,385].

Much research tends to focus on model training but what happens after a model is deployed is at least as important. Many high-stakes decisions are not made by algorithms alone and so investigating the ***interaction between humans and the machines*** that advise them is an important research objective. In many cases, human decision makers such as doctors and judges consider algorithmic predictions but must make the final decision themselves and understanding when and why they disagree is critical. This could improve fairness if the algorithm is flawed or not if the algorithm is correct but not trusted[386,387]. Furthermore, decision-making algorithms can create feedback loops and downstream effects which must be monitored and accounted for. In predictive policing, for example, the very act of increasing police presence in an area can cause the algorithm to continue suggesting crimes will be committed there[388].

Another big area of research is in assessing machine learning models which are not directly involved in decision-making but form a representation of the world which can be influenced by the same biases we have discussed. Large language models are an example of this. For instance, translation models have long struggled with assuming gender from context when translating between different languages. Models trained on large, unfiltered text can also learn to associate stereotypes which can amplify biases[389]. This in turn can affect downstream tasks such as sentiment analysis[390]. ***Representational bias*** can also happen in search results or algorithmic recommendations such as seeing mostly men when searching for images of CEOs[391].

Finally, the field of algorithmic fairness is deeply linked with ***explainable AI and causality research***. Models may feel unfair if there is no insight into how a decision was made or what aspects of the input led to a certain decision. Neural networks are often derided for being black boxes and navigating the competing desires for better accuracy and easier interpretability is a topic of much interest. Causal models could also play a role in ensuring fairness since they

384 Arnold, M., et al. (2019). FactSheets: Increasing trust in AI services through supplier's declarations of conformity. IBM Journal of Research and Development, 63(4/5), 6-1.
385 Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Iii, H. D., & Crawford, K. (2021). Datasheets for datasets. Communications of the ACM, 64(12), 86-92.
386 Green, B., & Chen, Y. (2019, January). Disparate interactions: An algorithm-in-the-loop analysis of fairness in risk assessments. In Proceedings of the conference on fairness, accountability, and transparency (pp. 90-99).
387 Kleinberg, J., Lakkaraju, H., Leskovec, J., Ludwig, J., & Mullainathan, S. (2018). Human decisions and machine predictions. The quarterly journal of economics, 133(1), 237-293.
388 Ensign, D., Friedler, S. A., Neville, S., Scheidegger, C., & Venkatasubramanian, S. (2018, January). Runaway feedback loops in predictive policing. In Conference on Fairness, Accountability and Transparency (pp. 160-171). PMLR.
389 Bolukbasi, T., Chang, K. W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? debiasing word embeddings. Advances in neural information processing systems, 29, 4349-4357.
390 Thompson, A. (2017, October 25). Google's Sentiment Analyzer Thinks Being Gay Is Bad. Vice. https://www.vice.com/en/article/j5jmj8/google-artificial-intelligence-bias
391 Kay, M., Matuszek, C., & Munson, S. A. (2015, April). Unequal representation and gender stereotypes in image search results for occupations. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (pp. 3819-3828).

attempt to relate features which reliably cause an outcome instead of just recognising patterns of correlation (see section 5.5 on causal AI). These models tend to be more easily explained since they work on logical principles and therefore have a great potential to be unbiased but they have yet to mature enough to be adopted at large scale.

Societal and media industry drivers

Vignette1: Detecting racial bias in targeted online advertising

Mark is the editor at an online magazine. One of the reporters has just published an article on a recent court case and Mark noticed an advertisement being shown on the article page was for a bail bond agency. Mark is African American. He asks some of his white colleagues what kind of ads they see with the article and one of them tells Mark he sees an ad linking him to a website which compiles the top ten law schools for criminal law. Mark wants to make sure the ads he runs are not targeting minority groups and reinforcing cycles of historical disadvantage. Was he being shown a bail bond ad specifically because he is black or was it a coincidence? He could ask more of his colleagues what kinds of ads they see but that would hardly provide a representative sample. After all, they all work in the same office and live in the same city. Mark wants to know overall, out of all the visitors to his site, what advertisements are being shown to each demographic group. He opens up the dashboard from the advertising company he uses and navigates to a screen showing fairness metrics for each ad shown on the website. The tool confirms that the bail bond agency is targeting minorities. Luckily, there is an option in the tool to equalise the serving rates for all ads in this category. Mark enables that option and checks back in a week to find that all demographic groups have an equal chance of being served any kind of legal ad.

Vignette 2: Detecting gender biases in content recommendation systems

Jennifer is a data scientist at a streaming video company. She wants to test out the recommendation algorithm for suggesting videos to watch next. Jennifer is worried that her daughter is constantly being shown makeup tutorials and hardly ever watches sports highlights anymore, even though she used to be a devoted tennis fan. She wonders if the recommendation algorithm has internalised certain gender stereotypes from its training data and is projecting them onto impressionable users. She decides to run an experiment: she creates two fake accounts, one as a young woman her daughter's age and one as a young man of the same age. She looks at the same content for one week on both and keeps track of what kinds of recommendations she gets. At the end, she tallies up the number of recommendations from each category and analyses her results. It appears that the algorithm did recommend slightly more sports videos but hardly ever recommended makeup videos without watching a few first to the male persona. Jennifer creates a report from her results and shows it to her manager. They agree to change the algorithm for young users to reduce representational assumptions which could cause self-fulfilling behaviour. Jennifer also talks to her daughter who says she looked up makeup tutorials on her own because she was interested and she enjoys them. She also tells Jennifer that she is still interested in tennis and would like to take lessons but she thought nothing exciting was happening since she wasn't being recommended any highlights.

Future trends for the media sector

Fairness must be considered in many media applications. The list below highlights a few examples likely to be relevant in the next few years:

Investigative journalism has revealed many of the shortcomings in algorithms cited above and will continue to play a key role in keeping the public informed and discovering unfair algorithms even as government regulation of AI increases. It is therefore integral that journalists and media professionals understand how machine learning algorithms can affect fairness, how to spot unfair algorithms, and what is being done about it.

Preventing discriminatory advertising actively is an important goal instead of just restricting demographic information from being used explicitly. In 2019, the U.S. housing department sued Facebook for discriminatory targeted advertising practices related to housing[392]. Facebook responded by preventing advertisers from using demographic information to target ads. Assessing the actual outputs for bias and preventing unintentional or subversive discrimination would be better.

Considering ***representational fairness*** can help prevent recommendation algorithms that perpetuate ***stereotypes*** and help break so-called ***filter bubbles***. Echo chambers in media are created when users are consistently shown the same type of content which can lead to corralling users into one of just a few different viewpoints when in reality, people are more multifaceted.

Diverse hiring is important in every field. With algorithmic resume filtering and candidate selection becoming widely used, it is paramount that we make sure these systems do not incorporate biases and, moreover, promote diversity.

Chatbots, automated text platforms usually used to provide support in lieu of a human, are becoming increasingly common and powerful. Automated textual interaction overall is a big trend and using large AI models such as GPT-3 without fully understanding them poses many risks. We have discussed how large language models can incorporate representational biases from data on which they are trained but more generally, as models become more general-purpose it becomes harder to test every use case and ensure no harmful text will be generated. Care must be taken when using models like this and understanding when and how they might break is crucial.

Furthermore, as ***generative AI*** gets better at creative tasks, we will begin to see it used to generate new stories such as movie, TV, and video game plots. AI has already been used in the

[392] HUD archives, HUD charges Facebook with housing discrimination over company's targeted advertising practice (2019): https://archives.hud.gov/news/2019/pr19-035.cfm

creative process of movie trailers[393] and advertisements[394]. Amidst growing calls for more diversity in film and TV, these models have the potential to introduce bias by rejecting ideas that do not cater to the majority and leaning into stereotypes if used naively.

"***Deepfakes***" are AI-generated images or videos which combine the action of one scene with a face from another. More generally, AI-assisted computer-generated imagery will become more prevalent and it is important to make sure this technology is used fairly and does not deny opportunities to diverse performers and that there is no gap in quality for performers of different appearances. These tools also bring up issues of consent and misinformation, related ethical issues to fairness, when used maliciously.

Visual intelligence tools or tools for photo/visual analysis have been widely discussed for their shortcomings in relation to recognising people and especially dark-skinned people. Great strides will continue to be made in remedying this situation but the greater question of whether the benefits of ubiquitous visual tracking are worthwhile is an ongoing debate. A slippery-slope argument says it can be hard to separate legitimate uses from objectionable overreach by, say, oppressive governments discriminating against minority groups. The first step before creating an algorithm should be anticipating and preventing its misuse.

Voice recognition and real-time voice transcription/captioning has been revolutionised by AI but it can still struggle understanding accents, dialects and less common languages. This technology can and is used extensively in media applications for accessibility but the benefits can be limited to native speakers of the language of the developers.

Abuse and hate speech detection on online platforms is a laudable ambition. The anonymity and platform scale mean automation is usually the only solution but the reliability of these systems to track harmful speech from and against different groups fairly has been questioned[395]. Robust and fair abuse detection is essential for the wellbeing of all.

Goals for next 10 or 20 years

The disparate effects of algorithms are considered from inception and all relevant stakeholders are included in the model building process. Decision-making algorithms are monitored for bias in production in real time to account for drift and feedback loops. Data and models are transparent and easy to understand; decisions are explainable and logical based on causal features. Representational models make guarantees about preventing the perpetuation of harmful stereotypes.

[393] IBM THINK Blog, IBM Research Takes Watson to Hollywood with the First "Cognitive Movie Trailer" (2016): https://www.ibm.com/blogs/think/2016/08/cognitive-movie-trailer/
[394] IBM THINK Blog, Lexus Europe Creates World's Most Intuitive Car Ad with IBM Watson (2018), https://www.ibm.com/blogs/think/2018/11/lexus-europe-creates-worlds-most-intuitive-car-ad-with-ibm-watson/
[395] Sap, M., Card, D., Gabriel, S., Choi, Y., & Smith, N. A. (2019, July). The risk of racial bias in hate speech detection. In Proceedings of the 57th annual meeting of the association for computational linguistics (pp. 1668-1678).

8 AI & data: building datasets and benchmarks for the future

The performance of AI models and the data they process during the training and testing phases of their development are inexorably linked. The recent surge in AI performance is undoubtedly influenced by the development of better models, better access to hardware processing resources, but also by the access to training and testing data created by the internet and social media. In this section, we analyse several aspects of the use of datasets in AI, presenting the status, challenges and trends for datasets and benchmarking competitions as well as ethical and legal aspects that relate to this domain.

8.1 AI datasets

***Contributors**: Mihai Gabriel Constantin (UPB), Mihai Dogariu (UPB)*

Current status

In the past few years, deep learning has become the de-facto approach in the multimedia field, spanning all types of applications and covering each major type of data, i.e., visual, textual, and audio. This has been strongly encouraged by the superior performances that deep learning methods have over classical machine learning algorithms. Many deep learning theories offer almost ideal results, but they are impossible to put into practice. Similarly, many deep learning models work well on specific tasks, but their creators cannot offer a sound explanation for choosing a specific setup. Therefore, there is still a reasonable amount of mystery concerning this field, where mathematical models lack practical implementation and applications lack a theoretical justification. It is a field where major breakthroughs on one of the two sides can propel the next breakthrough on the other side.

No matter what algorithm we design or implement, we are required to show its applicability through extensive validation. Each state-of-the-art method must prove its superiority when compared against existing research, so it is mandatory to have a common ground for these methods to establish a ranking. Traditionally, this has been done by reporting results on openly available datasets. The likes of MNIST[396], CIFAR-10[397], ImageNet[398], celebA[399], PASCAL VOC[400],

---

[396] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324

[397] Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images.

[398] Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009, June). ImageNet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition (pp. 248-255). IEEE.

[399] Liu, Z., Luo, P., Wang, X., & Tang, X. (2015). Deep learning face attributes in the wild. In Proceedings of the IEEE international conference on computer vision (pp. 3730-3738).

[400] Everingham, M., Van Gool, L., Williams, C. K., Winn, J., & Zisserman, A. (2010). The pascal visual object classes (voc) challenge. International journal of computer vision, 88(2), 303-338.

MS-COCO[401], SQuAD[402], GLUE[403], Penn Treebank[404], LibriSpeech[405], Universal Dependencies[406], VoxCeleb1[407], CheXpert[408] etc. have become representative benchmarks in their respective fields. They offer the premises for both training and validating algorithms, but they often lag behind the immense diversity that applications have reached. Figure 83 presents some samples from several of these datasets.

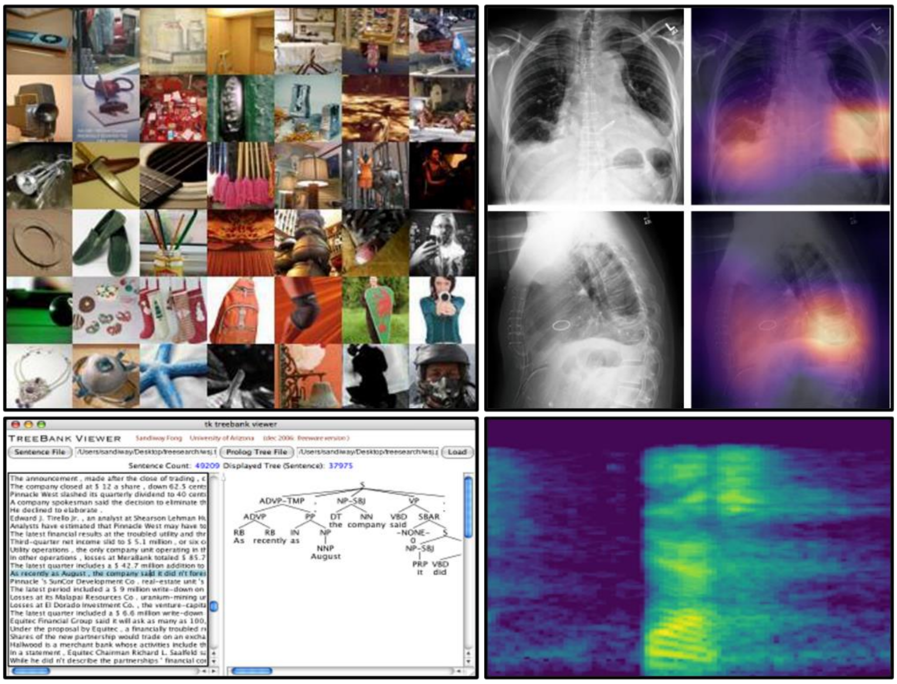


Figure 83: Samples extracted from large-scale public datasets. From left to right, top to bottom, samples from: ImageNet (images), CheXpert (chest X-rays), Penn Treebank (textual), LibriSpeech waveform (audio).

---

401 Lin, T. Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., ... & Zitnick, C. L. (2014, September). Microsoft coco: Common objects in context. In European conference on computer vision (pp. 740-755). Springer, Cham.

402 Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). Squad: 100,000+ questions for machine comprehension of text. arXiv preprint arXiv:1606.05250.

403 Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A multi-task benchmark and analysis platform for natural language understanding. arXiv preprint arXiv:1804.07461.

404 Marcus, M., Santorini, B., & Marcinkiewicz, M. A. (1993). Building a large annotated corpus of English: The Penn Treebank.

405 Panayotov, V., Chen, G., Povey, D., & Khudanpur, S. (2015, April). Librispeech: an asr corpus based on public domain audio books. In 2015 IEEE international conference on acoustics, speech and signal processing (ICASSP) (pp. 5206-5210). IEEE.

406 Nivre, J., De Marneffe, M. C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C. D., ... & Zeman, D. (2016, May). Universal dependencies v1: A multilingual treebank collection. In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16) (pp. 1659-1666).

407 Nagrani, A., Chung, J. S., & Zisserman, A. (2017). Voxceleb: a large-scale speaker identification dataset. arXiv preprint arXiv:1706.08612.

408 Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haghgoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., & Ng, A. Y. (2019, July). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 590-597).

Research challenges

Research in the AI field faces several challenges when it comes to working with datasets. Below, we list the most important ones:

Low dataset diversity: Most AI algorithms have a limited applicability due to the fact that there are no large and diverse enough datasets that correspond to their needs. This leads to training done on other datasets than what is specifically needed for the task at hand. For example, there are numerous algorithms that are pre-trained on ImageNet, even though they do not aim to classify images. One could argue that ImageNet is the most complete large-scale dataset which helps best initialise a network's weights, but it is far from perfect for each computer vision task. In response to this, researchers might opt to create their own datasets, which leads to the following issue.

Complex data gathering process: Creating a relevant dataset involves a great deal of effort from the involved team. Gathering the dataset samples is just the start of a tedious process and, even at this point, critical problems may arise. For example, satellite images are not easy to acquire since it implies tremendous costs to launch a satellite for this purpose. Therefore, researchers have to deal with whatever resource is available for them. Moreover, even if researchers have hypothetical access to unlimited samples, they still need to annotate them. An accurate annotation of the dataset is critical for the outcome of any application. However, the human resources needed to accomplish such a task is expensive because it requires many persons to perform this process and they also need to be accurately trained. Moreover, the annotator's subjectivity may also interfere with the goodness of the annotations. After such a lengthy process, a large part of the institutions or groups that gather such datasets may not be too keen to share them with others, especially if it gives them a competitive edge, which leads to the next issue.

Closed datasets: The high costs of creating a good dataset may deter researchers from freely sharing their work. This is, most often, the case for companies involved in research that want to ensure an advantage over their competitors. It is by all means understandable that they choose to do so, since this contributes to their source of revenue, but the research results that they report are difficult to validate by external researchers. Even more, a large number of datasets are inaccessible due to privacy concerns, which raises the following point.

GDPR concerns: With the introduction of the GDPR regulations in 2018, most processing that involves personal data has hit a wall in development. While these regulations came to the aid of individuals, they have seriously hindered the gathering of data for numerous datasets. Some examples include face detection or speaker verification. Identity obfuscation becomes, this way, a very appealing field that can help the creation of large-scale datasets.

Bias issues: AI algorithms are known for offering objective results from the decision point of view. However, there is no warranty that this objectivity transfers to the system's global decision philosophy, especially since any bias in the training dataset will be visible in the output. This will skew the system's capabilities in favor of the most frequent/prominent samples of the dataset, leading to falsely accurate results (see section 7.4 on AI fairness). For example, Shankar et al.[409] examined two large-scale datasets, Open Images and ImageNet, and discovered that these datasets are highly biased towards the countries where they were gathered. They are highly US and Euro-centric, leaving other geographical areas, such as the entire African continent, severely under-represented. The same authors present a pie-chart of the distribution of geographically identifiable images in these two datasets, shown in Figure 84.

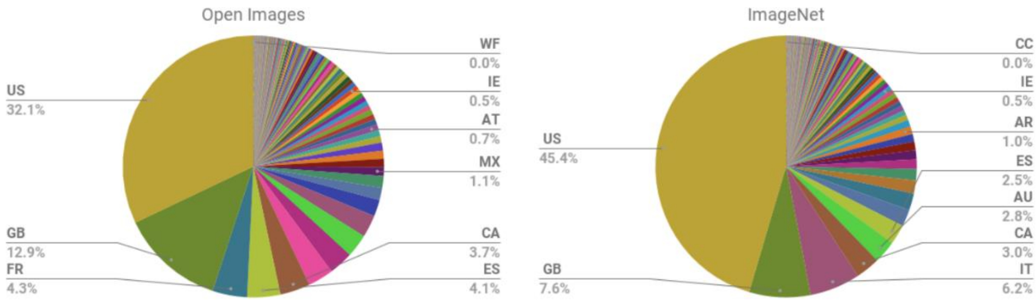


Figure 84: Fraction of Open Images and ImageNet images from each country [Shankar et al, 2017][409].

Proper benchmarking: Large-scale datasets often come with a validation protocol that researchers 'should' follow. Not too rare is the case when researchers will force different training + validation scenarios such that their proposed approach will top the existing state-of-the-art. The current state of worldwide research emphasises outperforming the state-of-the-art to the extent that researchers are tempted to alter the validation process just so they obtain results better than the state-of-the-art. This is usually noticed when trying to replicate the experiments, and end up with unexpected results when following the dataset's official validation protocol.

Future trends for the media sector

AI algorithms evolve at an astonishing pace. It is crucial that these algorithms are trained and validated adequately on large-scale datasets. These datasets lay at the very heart of all learning algorithms and they play a vital role. With people relying more and more on technology, especially on AI systems, we will start feeling the flaws of these datasets in our lives. The media sector can help in this regard since they have access to or are the owners of large data archives, which they could share partly or fully with the research community. Researchers would benefit tremendously from such datasets since they are at least weakly annotated (having a title, short description, metadata available etc.). These data collections also have the advantage of spanning large time periods, so a certain chronological evolution of its constituent samples can be observed, e.g. how fashion changed in time, how the media-specific terms and phrases differ now from the past, how image quality improved, etc.

409 Shankar, S., Halpern, Y., Breck, E., Atwood, J., Wilson, J., & Sculley, D. (2017). No classification without representation: Assessing geodiversity issues in open data sets for the developing world. arXiv preprint arXiv:1711.08536.

The multimedia research community is directly involved in the development of most smart applications that are pushed into production, so it should also take responsibility for its drawbacks. In the future, there will be a great deal of emphasis placed on individual privacy and AI invasiveness. Therefore, it is the media sector's responsibility to ensure that any attempt to violate privacy rights will be identified and reported. Investigative journalism can have a major impact on holding accountable companies or states that take invasive actions towards their clients or citizens.

Lastly, tech giant companies will compete for the best large-scale datasets, since they are amongst the few organisations that hold the necessary resources to carry on such an effort (e.g. Microsoft's Common Objects in COntext dataset, Google Open Images, Habitat - Matterport 3D, etc.). There are numerous start-ups, spin-offs and independent businesses that are being born every day and a small part of them become visible with innovative products. This draws the attention of tech giant competitors which might buy and integrate them into their portfolio, e.g., WhatsApp acquisition by Facebook. Instead of focusing on the record-breaking transactions by these tech giants, the media sector should try to emphasise more the role that smaller companies have in the entire AI tech ecosystem. This way, the effort of rising visionaries will not be diluted under the umbrella of big companies.

Goals for next 10 or 20 years

The future of AI applications is changing with each passing day. Applications that seemed to be unattainable during our lifetime are now already obsolete and the future holds as many new exciting applications as the human imagination can fathom. As previously stated, datasets play a vital role in designing performant algorithms. Therefore, to overcome some of the current setbacks, AI datasets should deal with several aspects:

Open data: The first and most important characteristic of AI datasets is that they should be open for research. The biggest advances in any field came when research was shared with other peers. There are several attempts at the moment to centralise access information on all large-scale multimedia datasets, but the future might offer a centralised way of creating, uploading and sharing datasets among researchers. Data openness (under particular legal constraints) is a must for the future of AI.

Automatic annotations: Simply put, humans do not have the ability to annotate media content at the same rate that it is being produced. Automatic annotations have the potential to solve the limited availability that large-scale datasets have.

Synthetic data generation: Another way of solving data scarcity is to generate completely synthetic samples that resemble or complete the characteristics of the original dataset. This is already a promising approach with the rise of GANs.

Privacy: datasets should be built with the idea of private personal information in mind. This problem seems to become more and more prevalent in every application since reports about consumer applications violating individual privacy are surfacing every day.

8.2 AI benchmarks

Contributors*: Mihai Gabriel Constantin (UPB), Mihai Dogariu (UPB)*

Current status

The latest trends in AI show an increased attention for the creation and wide adoption of benchmarking competitions. Media data has been constantly featured in different conferences that are dedicated to the creation of common benchmarking systems[410,411,412], that integrate common subject matter definition, data, train-development-test splits, metrics, annotations and pre-processed auxiliary data.

While organisers and participants create and use common metrics in understanding system performance, an interesting development is that benchmark organisers are also interested in the discovery of general trends that go beyond just ranking the systems according to their performance. These trends may be defined by ideas, approaches, or parts of approaches that not only tend to influence the performance of methods, but that also shed some light into the studied concept itself. To this effect, the MediaEval[412] website declares interest in studying the "*Quest for Insight: List several research questions related to the challenge, which the participants can strive to answer in order to go beyond just looking at evaluation metrics.*", while the ImageCLEF[411] website declares that "*The main goal is not to win the competition but compare techniques based on the same data, so everyone can learn from the results. Everyone who has been to a workshop can tell that the discussions at the posters are always very vivid and many approaches that work or do not are discussed. It is a chance for everyone to learn from each other*".

The current status of AI benchmarking systems is also dictated by the emergence and adoption of Evaluation as a Service (EaaS) systems. EaaS is defined as a cloud computing based architecture that can be used for assessing or evaluating a series of systems, by providing the necessary tools for system evaluation[413,414]. This type of approach attempts to increase the ***reliability, trustworthiness and reproducibility*** of the AI methods, as well as ensure a fair chance for each participant to the benchmarking competitions. While in no way an exhaustive list, just a few of the most important and popular EaaS platforms are Kaggle[415], AIcrowd[416] or Codalab[417]. Most of them are open source, allowing their integration in other projects, as well as changes and improvements brought by interested parties, and present many useful functions. Some examples would be:

---

410 CLEF: http://clef2021.clef-initiative.eu/
411 ImageCLEF: https://www.imageclef.org/
412 MediaEval: https://multimediaeval.github.io/
413 Hanbury, A., et al. (2015). Evaluation-as-a-service: overview and outlook. arXiv preprint arXiv:1512.07454.
414 Wasik, S., Antczak, M., Badura, J., & Laskowski, A. (2018). Evaluation as a Service architecture and crowdsourced problems solving implemented in Optil. io platform. arXiv preprint arXiv:1807.06002.
415 Kaggle: https://www.kaggle.com/
416 AIcrowd: https://www.aicrowd.com/
417 Codalab: https://codalab.org/

- ***Leaderboards and scoring***: allowing organisers to create and use their own metrics and methods of scoring participant runs and displaying their performances.
- ***Sub-tasks, multi-phase, and scheduling***: allowing organisers to create multiple sub-tasks with different targets, create multiple phases for the tasks and schedule data publication accordingly.
- ***Computing workers and containerisation***: allowing organisers to create worker cloud-based virtual stations that can be used for the task, and allowing participants to submit containerised versions of their method, thus supporting reproducibility of the results.

Some examples of previous, current or upcoming AI benchmarking campaigns published by the AI4Media project include the following:

- **Interestingness10k**[418] provides participants with image and video samples extracted from Hollywood-like movies, annotated for visual interestingness. The organisers also provide common metrics, data splits, pre-extracted features, as well as an in-depth analysis of previously used methods for this set of data.
- **MediaEval Predicting Media Memorability** 2020[419] and 2021.[420] This task addresses short- and long-term memorability for videos. The authors also provide a common set of metrics, data splits and pre-extracted features. Also, for the 2021 edition of the benchmarking campaign, two different sets of data are provided, and a generalisation and EEG-based subtask are available for interested participants.
- **ImageCLEFaware** 2021[421] and 2022[422] proposes a task where participants are asked to infer the effect of media content sharing on several real-world situations, such as: asking for a bank loan, getting an accommodation, searching for a job as a waiter/waitress, and looking for a job in IT. The organisers provide a common set of metrics, anonymised user profile data, and automatically extracted predictors.
- **ImageCLEFfusion** 2022[423] proposes a benchmarking task for comparing the performance of late fusion or ensembling systems, applied to two media processing tasks, namely media interestingness and image search result diversification. The organisers provide a common set of metrics, data splits, as well as the outputs from a pre-processed set of inducers that will be used by participants as inputs for their fusion systems.

Research challenges

The use of AI benchmarking competitions has increased over the last 10 years, and popular competitions like ILSVRC[424] ended up defining some of the most impactful trends in AI over this timeframe. Some of the most important challenges in this domain can be summed up as follows:

[418] Interestingness10k: https://www.interdigital.com/data_sets/interestingness-dataset
[419] MediaEval PMM 2020: https://multimediaeval.github.io/editions/2020/tasks/memorability/
[420] MediaEval PMM 2021: https://multimediaeval.github.io/editions/2021/tasks/memorability/
[421] ImageCLEFaware 2021: https://www.imageclef.org/2021/aware
[422] ImageCLEFaware 2022: https://www.imageclef.org/2022/aware
[423] ImageCLEFfusion 2022: https://www.imageclef.org/2022/fusion
[424] ImageNet Large Scale Visual Recognition Challenge (ILSVRC): https://www.image-net.org/challenges/LSVRC/

Reproducibility: It is of utmost importance to ensure that the results submitted by participants can be reproduced according to the requirements for the testing set data. This can be ensured by the integration of containerisation or open source submission that would allow participants to submit their proposed model and allow the community to check their results.

Trustworthiness: Given that many of the most popular AI methods and models are created or gain their popularity during AI benchmarking campaigns, it is important to ensure their trustworthiness, including concepts like explainability, robustness, fairness, transparency and privacy. It is important for future competitions to ensure that these aspects are thoroughly treated by all the participants of the competitions.

Continuity: While it is important to have a clear schedule and draw the main conclusions during the timeframe of the competition, it may also be a good idea to continue and keep the resources open after the competition is done. In this way, organisers can ensure that future users of the dataset attached to the competition respect the same conditions as all the other participants.

Fair access: The growth of AI resource availability helped a lot of researchers test their work; however, there are cases where hardware resources are not available, either through students with low access to GPU computing capabilities or cases when participating in the competition is not a priority at that time.

Efficiency and green computing: This represents another new dimension for measuring the performance of AI systems. Along with measures of trustworthiness and traditional performance metrics, an evaluation of how impactful training and running a system is on the environment creates a better and more accurate picture for system integrators, private companies and authorities with regards to the real performance of AI methods and models.

Future trends for the media sector

Future trends will undoubtedly take into account some of the current challenges and problems currently faced by this domain. The media sector in particular can be helped by the contributions and cooperation of large media companies, social media platforms, either through sponsoring, endorsing or contributing with data to the AI benchmarking competitions.

One important aspect with regards to media companies that are interested in organising benchmarking competitions is represented by data sharing anonymity. While current regulations and industry standards provide guidelines for user data protection, future developments may present a great opportunity for the introduction of fully or partly synthetic datasets, therefore creating a very powerful layer of anonymity.

The media industry will also represent an important factor in setting the definition and use case scenarios for the target concepts. One such current example is the Interestingness10k dataset, where the industry co-organiser proposes a use case scenario that is important from a practical standpoint for their applications or business model, namely helping content creator professionals select the most interesting frames or video excerpts from their movies in order to list them on Video-on-Demand platforms.

Also, it is important to note the responsibility of AI benchmark organisers to ensure that the competitions take into account all the aforementioned aspects. In this regard, there is a clear need for more emphasis on the trustworthiness aspect of AI models.

Goals for next 10 or 20 years

The future of AI benchmark competitions will be defined by the future of EaaS technologies, as these aspects will be closely interconnected and applying EaaS technologies to AI benchmarks will ensure a fair and open competition for all participants. While cloud-based computing has been developed for a long time, the next period may see an increase in its presence in computer vision, natural language processing, and artificial intelligence in general, as the technology may become more ubiquitous and may see a decrease in costs. Therefore, it may become possible for competitions to become more complex, either by integrating certain trustworthiness metrics as a supplementary dimension of performance or by exploring several concepts at the same time and considering their correlations. Also, cloud-based computing power may increase to shared GPU computing that can target fair chances of access to participation and may increase the rate of reproducible methods. From a software standpoint, it will be important for organisers to create open APIs that help participants submit their models in a unified way that allows for countless variations with regards to the software packages they employ, but ensure a standard input and output, therefore easing reproducibility.

8.3 Ethical and legal aspects of availability of quality data for AI research

***Contributors**: Lidia Dutkiewicz (KUL), Noémie Krack (KUL), Emine Ozge Yildirim (KUL)*

Current status

Considerable amounts of training and testing data are necessary for research and development of AI especially for Machine Learning techniques. The higher the quality of the training data, the better the system outputs will be. AI researchers are therefore looking for vast datasets in order to produce reliable and accurate outputs. For training AI models, data such as CommonCrawl[425], used for training large language models, ImageNet[426] for object recognition, or MS COCO[427] for computer vision tasks are employed. Not all data contained in these massive datasets constitute personal data and hence are not covered by the data protection legal framework; however, parts of them may contain personal data and even special categories of personal data.

[425] Common Crawl: https://commoncrawl.org/
[426] ImageNet: https://www.image-net.org/
[427] COCO dataset: https://cocodataset.org/#home

The literature has been highlighting the issues of ***quality and representation*** in training data as part of publicly available datasets and databases[428]. Pornography, stereotypes, racist and other problematic contents were found to be part of datasets[429]. When used in AI research, the biased, incorrect training data will replicate or exacerbate these negative features in the AI system outputs[430]. This can be illustrated by the ***'garbage in, garbage out' principle***, opening the door to human rights violation and discrimination[431].

Data protection in research and its legal issues have been the subject of study by various scholars that we recommend reading[432,433,434,435,436,437]. Several challenges have been identified including:

- the lack of legal basis to process personal data in the publicly available datasets,
- the difficulty to comply with data subjects' rights,
- the difficulty to comply with technical and organisational safeguards to process data in a data protection-compliant way.

In addition to the accuracy problems linked with low quality datasets, ethical considerations arise including research quality, legitimation of the use of biased datasets. This will not only harm the quality of the research outputs, but also its further use and the more general academic reputation. Some may remember the MegaFace[438] database. Megaface was a publicly available and widely used benchmark datasets in AI research set up by the University of Washington. Thanks to researchers, journalists and activists, the alarm was triggered regarding the issues present in these datasets in respect to the right to privacy and other human rights[433]. However, repeated scandals about training data available and AI research will hinder public trust in science.

---

428 Raji, Inioluwa Deborah, Timnit Gebru, Margaret Mitchell, Joy Buolamwini, Joonseok Lee, and Emily Denton. 2020. 'Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing'. In , 145–51. https://doi.org/10.1145/3375627.3375820.
429 Birhane, Abeba, Vinay Uday Prabhu, and Emmanuel Kahembwe. 2021. 'Multimodal Datasets: Misogyny, Pornography, and Malignant Stereotypes'. ArXiv:2110.01963 [Cs], October. http://arxiv.org/abs/2110.01963.
430 Osoba, Osonde A, and William Welser IV. 2017. 'An Intelligence in Our Image: The Risks of Bias and Errors in Artificial Intelligence'. Rand Corporation. https://www.rand.org/content/dam/rand/pubs/research_reports/RR1700/RR1744/RAND_RR1744.pdf
431 Richardson, Rashida, Jason Schultz, and Kate Crawford. 2019. 'Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, and Justice'. New York University Law Review 94: 42.
432 Ducato, Rossana. 2020. 'Data Protection, Scientific Research, and the Role of Information'. Computer Law & Security Review 37: 105412.
433 Jasserand, Catherine. 2018. 'Massive Facial Databases and the GDPR: The New Data Protection Rules Applicable to Research'. In Data Protection and Privacy: The Internet of Bodies, 169–88. Hart Publishing/Bloomsbury Publishing Plc.
434 Weller, Katrin, and Katharina E Kinder-Kurlanda. 2016. 'A Manifesto for Data Sharing in Social Media Research'. In , 166–72.
435 Barfield, Woodrow, and Ugo Pagallo. 2018. Research Handbook on the Law of Artificial Intelligence. Edward Elgar Publishing.
436 Barocas, S., and A. Selbst. 2016. 'Big Data's Disparate Impact'. California Law Review 104 (671).
437 De Bruyne, Jan, and Cedric Vanleenhove, eds. 2021. Artificial Intelligence and the Law. Intersentia.
438 Megaface database: http://megaface.cs.washington.edu/

Given the above elements, improving access to relevant and sufficient data for AI research is needed more than ever. Complementary initiatives contributing to the fight against problematic data proliferation and use are equally needed. A recent study[439] put forward measures which would improve the situation. This includes making ethically salient information about datasets clear and accessible, active stewardship of the data and its uses, employing ethics review procedures to promote responsible data uses, and the advance review of datasets and publications.

The scholar community is also willing to address these issues and many journals announced extra ethics checks on AI research to be published. We can name the Conference and Workshop on Neural Information Processing Systems (abbreviated as NeurIPS and formerly NIPS), where papers could be rejected due to ethical and legal doubts associated with the data used[440].

The situation is not easy for AI researchers as doing AI research means coping with the fundamental tension between data protection (including data minimisation) and the vast amount of data needed for meeting accuracy and quality requirements. Furthermore, the diverse and vague legal frameworks are not helping to provide clear guidance to AI research[441].

All of these challenges can constitute a barrier to AI research harming the research impact, the publishing opportunities and society's trust in academia and research in general. Trust is essential in scientific research.

Possible ways forward

Availability of quality data is a core part of European technology policy discussions. Already in 2018, the European Council acknowledged that '*high-quality data are essential for the development of AI*'[442]. A year later, the HLEG Ethics Guidelines for Trustworthy AI[443] referred in particular to ***data governance*** as one of the requirements for AI systems, highlighting the impact of biases in training datasets. In 2020, the European Commission released its White Paper on AI[444], which underlined how ***data availability*** was essential for training AI systems and how data quantity and quality were key components for building trustworthy and unbiased AI systems.

[439] Peng, Kenny, Arunesh Mathur, and Arvind Narayanan. 2021. 'Mitigating Dataset Harms Requires Stewardship: Lessons from 1000 Papers'. ArXiv Preprint ArXiv:2108.02922.
[440] Beygelzimer, Alina, Yann Dauphin, Percy Liang, and Jennifer Wortman Vaughan. 2021. 'Introducing the NeurIPS 2021 Paper Checklist'. Medium (blog). 26 March 2021. https://neuripsconf.medium.com/introducing-the-neurips-2021-paper-checklist-3220d6df500
[441] Rogers, Anna, Tim Baldwin, and Kobi Leins. 2021. 'Just What Do You Think You're Doing, Dave?'A Checklist for Responsible Data Use in NLP'. ArXiv Preprint ArXiv:2109.06598.
[442] European Council. 2018. 'European Council Meeting of 28 June 2018 – Conclusions (EUCO 9/18)'. https://www.consilium.europa.eu/media/35936/28-euco-final-conclusions-en.pdf.
[443] High-Level Expert Group on Artificial Intelligence. 2019. 'Ethics Guidelines for Trustworthy AI'. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.
[444] European Commission. 2020. 'White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (COM(2020) 65, Final)'. Brussels. https://ec.europa.eu/info/sites/info/files/commission-white-paper-artificialintelligence-feb2020_en.pdf

To meet the quality requirement, the EU policy instruments provide several measures aiming to solve the challenge relating to the lack of available data for the EU digital transformation:

- The ***European data strategy***[445] puts forward firstly EU-wide common, interoperable data spaces in strategic sectors as a solution to this lack of available data. Set up by the Commission, they would provide trustful, accountable and non-discriminatory access to high-quality data for the training, validation and testing of AI systems. Secondly, a horizontal governance framework for data access and use is said to '*facilitate decisions on which data can be used for scientific research purposes in a manner compliant with the GDPR'*.
- The ***Data Governance Act***[446] proposal puts forward a new interesting concept called *data altruism,* according to which data can be made available for purposes of 'general interest'.
- The ***EU AI Act***[447] proposal provides that for the development of high-risk AI systems, certain entities, such as digital innovation hubs, testing experimentation facilities and researchers, '*should be able to access and use high-quality datasets within their respective fields of activities*' (Recital 45).

Recent policy and legal initiatives show how importance on the availability of quality data has been acknowledged by policymakers. It remains to be seen how this will materialise tangibly for AI researchers. The research community is in need of effective tools and clear guidance to operate legally. Despite these initiatives, a wide binding data access framework for AI research would be more than welcome to enable researchers to access data in a harmonised and legally-compliant way.

[445] European Commission. 2020. Communication on a European Strategy for Data. https://digital-strategy.ec.europa.eu/en/policies/strategy-data
[446] European Commission. 2020. Data Governance Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52020PC0767
[447] European Commission. 2020. Artificial Intelligence Act. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206

9 AI applications and solutions for the media sector: imagining the future of next-gen media

In this section, we focus on the use of AI in different media industry sectors, including news, social media, film/TV, games, music and publishing. We also explore the use of AI to address critical global phenomena such as disinformation. In addition, we examine the use of AI to enhance online political debate. Finally, we discuss how AI can help the study of media itself in the form of AI-enabled social science tools.

Each subsection is dedicated to a different media sector or topic, following the same format that was adopted for section 5 (see the introductory paragraph of that section), attempting to examine the current status, applications, challenges, and future trends for the use of AI in the media industry.

9.1 AI for citizen participation and democracy

***Contributors**: Dario Garcia Gasulla (BSC)*

Current status

During the last couple of decades, political and social communication has been transformed by digital technologies, and particularly, social media. Massively populated online platforms have changed the dissemination landscape, enabling huge audiences of anonymous citizens, and reciprocally, offering unlimited and unverified sources of information to all. While this is a huge step in the democratisation of information, opinion and public attention, it entails novel challenges that must be dealt with.

Among the challenges of modern communication is the economy of attention. Citizens have a limited capacity for consuming information, creating a coveted (both by producers and consumers) market of audiences. Citizens wish to make a more efficient, productive and controlled use of their attention capacity, being increasingly aware of the limits of their own attention. Meanwhile content producers wish to gather as much public attention as possible, often with little regard for ethical, reliability and quality issues.

The rise of the attention economy plays a big role in current democracies, and the participation of citizens in it. People are more informed (in terms of volume, but not necessarily in terms of quality) and empowered than ever, having powerful tools at their disposal for defining and disseminating their opinion. However, citizens are also subject to an unprecedented amount of information bombardment, often of poor quality, designed mainly for the purpose of gathering attention. Figure 85 shows a broad categorisation of the threats individuals face through misinformation, echo chambers, filter bubbles and others. This is our target for the protection of democracy and social trust.

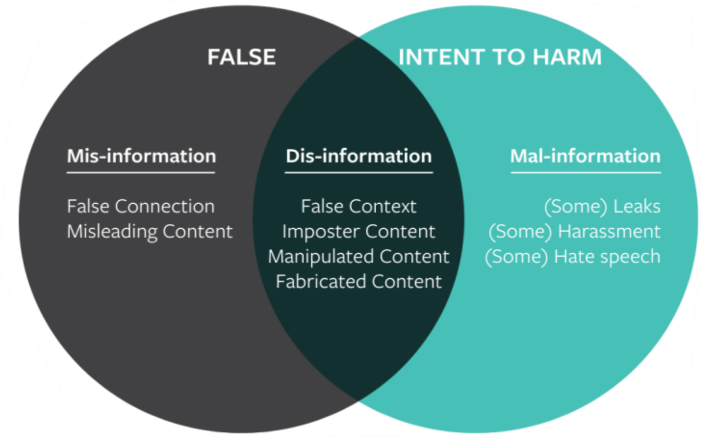


Figure 85: Categorisation of information disorders.[448]

Research challenges

A tricky balance in current democratic and participatory policies has to be maintained between fundamental rights, like free speech and freedom of opinion, and the use of tools which allow users to navigate the sea of information (reliable or not) that we live in. This entails many different aspects regarding quality (is information of good quality?), content (is content adequate for the audience?), means (is information presented in an acceptable way?), and variety (is available information representative of the whole spectrum?). There are many possible dimensions for this assessment, as shown in Table 4. To cover them all, that is, to provide a comprehensive view of the problem, we must consider it from many different perspectives, including:

- ***Sentiment analysis***: In which direction is this content pushing?
- ***Polarity***: How polarised is this content?
- ***Offensiveness***: Does the content contain abusing or degrading statements?
- ***Botness***: Is this content produced by a human in a non-automated manner?
- ***Propaganda***: Is the purpose of the content to push an underlying agenda?
- ***Ephemerality***: Is the content produced in a prolonged discussion or in a burst of one way information?
- ***Fallacious argumentation***: Is the content using mechanisms of misdirection?
- ***Synthetic manipulation***: Has the content been tampered with or altered in undisclosed ways?

Table 4: Dimensions of disinformation[449]

	Motive

[448] Image source: Wardle C, Derakhshan H. Thinking about 'information disorder': formats of misinformation, disinformation, and mal-information. Ireton, Cherilyn; Posetti, Julie. Journalism, 'fake news'& disinformation. Paris: Unesco. 2018:43-54.
[449] Table adapted from: Kapantai E, Christopoulou A, Berberidis C, Peristeras V. A systematic literature review on disinformation: Toward a unified taxonomical framework. New media & society. 2021 May; 23(5):1301-26.

Dimensions/ measurement	Profit	Ideological	Psychological	Unclear
Clickbait	✓		✓	
Conspiracy theories		✓	✓	
Fabrication	✓			
Misleading connection	✓			
Hoax	✓			
Biased or one-sided	✓			
Imposter	✓			
Pseudoscience	✓		✓	
Rumors	✓			
Fake news	✓			
Trolling	✓			

Considering all these topics at the same time is a considerable challenge, as is integrating them. However, this is a necessary effort, given the relevance of the underlying purpose (enabling a stronger and more reliable participation of citizens in democracy, and strengthening democracy's foundations). Coherently, each of these topics must be addressed, first separately, and then in conjunction. The behaviour of all these discussion characteristics may be correlated (some indicators may most frequently go together) and may be complementary, providing a holistic view of the nature of information. All these relations must be explored, analysed and exploited in the next few years, for the sake of citizens, their interactions with the digital world, and democracy itself.

Societal and media industry drivers

Vignette: Analysis and validation of information from online debates for journalistic reporting

Camila is a journalist that reports on social hot topics, using online discussions as a guideline of what is generating more interest and controversy. In her search for what is going on, Camila reviews social network activity (both content and users) and annotates what seems most relevant. Then, for each topic selected, Camila performs an in depth search on social media, looking for interesting information sources, online profiles of reference, and general argumentative points used around the issue.

Before transforming all that into a news piece that will be shared through official channels, Camila verifies several aspects of the gathered data. She uses automated tools to identify properties such as the polarity, offensiveness, botness, ephemerality and others, related both with the content and users participating and generating it. Based on this analysis, she decides to put aside all content, arguments and interactions produced by potentially malicious agents (e.g., detecting bots, synthetic manipulation), as these have to go through a specific double verification channel. Then, she explores how representative of society is the content she has gathered (e.g., measuring the polarity, sentiment analysis), and identifies sides of the story which are not found in the available narrative. Camila is a very thorough journalist, who wants to be as objective as possible. For this, she looks for agents with economic interests in the topics of argumentation (e.g., identifying propaganda), as this will allow her to add disclaimers and labels on her report regarding potential conflicts of interest. Camila also makes an effort to make sure that all content is appropriate for all audiences (e.g., measuring offensiveness), and writes her work focusing on those parts of the conversation which exhibit a higher overall quality (e.g., tracking ephemerality and fallacious argumentation).

Once she is finished, Camila is proud of her work, and assured that it will contribute to produce better informed citizens, reducing the amount of information noise being shared in current society.

Future trends for the media sector

The role of journalists is shifting. From the investigation, content generation and creation of opinion duties typical of the 20th century, to assuming duties on information curation, summarisation, and point of view weighting. Developing and assessing tools to support these new responsibilities is a priority. As is the familiarisation of all actors (media creators, professional or not, and media consumers) with the existing risks and the tools available to handle them.

Future trends in opinion mining, propaganda identification, fallacious argumentation detection, etc. are likely to focus on the means instead of the contents. Given the complexity of objectively measuring the degree of truthness in a given discourse, the field is pivoting towards characterising the processes, means and practices that most easily result in unreliable, misleading or manipulated content.

While these tools are being designed with novel communication means in mind (mostly online), they are also applicable to traditional media. Polarisation in written media, propaganda in films and TV shows, fallacies in cartoons and strips, offensive content in videogames, among others. We expect these other applications to be prioritised as tools become more proficient, and their benefits more obvious.

Goals for next 10 or 20 years

In the medium to long term, we expect AI tools to be fully integrated into all social media sites. Citizens and journalists will consume information through them, and the access to raw data feeds will be rare and discouraged. For doing so, a high level of transparency, understanding and trust between AI methods and users will be necessary.

These tools will provide warning signs, information of reliability, suggestions for a deeper and/or wider understanding of the domain, and other functionalities that empower citizens, and reduce their vulnerability in front of the big data.

9.2 AI for news production

***Contributors**: Sabino Metta (RAI), Alberto Messina (RAI)*

Current status

News is more than just a tangible product. It is a public good that contributes to the pillars of democracy. The availability of news allows citizens to be informed about issues that concern them so that they can make better decisions. So, it is fundamental that news is produced professionally and ethically. On the other hand, the digitalisation of media is supposed to increase the tension between the production of news as a public good vs. its delivery as a commodity. Nowadays, news is made readily available and accessible on multiple sources: online, radio, television, social networks, etc. Information is continuously produced by large groups of professional journalists and communication experts. It is also ourselves and our own network of people that continuously produce and propagate information simply by sharing them. In this media ecosystem the "time" variable plays a crucial role. Journalists are focusing on producing news quicker with respect to reporting more in-depth news and the entire news production chain is stressed to work at high speed.

Figure 86: A Rai News 24 reporter reporting from London.

Technology in the newsroom has helped to "speed up" news production and distribution, including personalisation, marketing, finding audiences, understanding user behaviour, monetisation/subscriptions[450]. Technological advancements are aiding journalists to deal with the pressure in the newsrooms. By using light devices (mobiles, tablets, etc.) journalists can record and report on news from wherever they are and easily upload items onto news sites as well as social media. Tools allow journalists to search and retrieve specific contents stored in newsroom archives thus enriching their stories. The speed of the internet enables journalists to find and get data without having to leave the newsroom. Furthermore, the internet allows journalists to gather and analyze possible comments provided from end users thus inspiring new investigations. In addition, in the age of digital and social media, visual communication has sky-rocketed. Images are powerful, we know. Visuals quickly transmit information to our brain and trigger our emotions thus stimulating our engagement. This can be particularly difficult for TV news production where the storytelling is based on the availability of images and videos. When a chronicle fact occurs, live footage allows to tell and to show what has been happening. In case of unexpected events (like natural disasters: floods, earthquakes, pandemic, etc.), newsroom's

[450] C. Beckett. "New powers, new responsibilities. A global survey of journalism and artificial intelligence". November 18th 2019, JournalismAI | Research, London School of Economics and Political Science.

correspondents need to readily publish and update fresh news in a very short timeframe (Figure 86).

To manage very low bandwidth video transmission, ***AI technologies can increase the efficiency of video compression***. Algorithms calculate bitrate to optimise bandwidth usage while maintaining an appropriate level of quality.

Audiovisual archives can enrich and empower storytelling. When a flood (or any other natural disaster) occurs, audiovisual archives can be used for documenting environmental and urban conditions of a location before the event occurred. In this way, damages caused by the flood can be highlighted by images before and after the disaster has occurred. Currently, media organisations face several issues regarding access to their archives: different and time-changing A/V formats, video content rights, different definitions (standard, FullHD, 4K, etc.), several types of compression (MPEG2, MPEG4, AVC, etc.), metadata, wrappers (MXF, Quicktime), large A/V storages (petabytes). Content Management Systems oversee making video content findable, referenceable and reusable.

On the other hand, ***AI technologies can enhance video archives***. These technologies can efficiently automate different tasks: metadata enrichment, splitting of long videos, object and scene identification, facial recognition, audio transcription, on-screen text extraction, etc. Machine learning technologies can automate the generation of searchable metadata tags by timecode (about actions, places and things identifiable within the content, etc.) thus enhancing search and discovery tasks in content libraries. In addition to the opportunity to properly document occurring events, the extraction of generic scenes within video segments (e.g. a rainy day, a landslide in the mountains, etc.) can be opportunistically exploited for evoking ad-hoc context and enhancing storytelling[451,452].

Recent developments in deep learning models are revolutionising the traditional workflow for video creation. In 2018, China's state news agency Xinhua developed AI anchors through machine learning technologies. Voice, facial movements, and gestures of real-life broadcasters have been simulated[453]. Currently, there exist several AI video generator tools that provide virtual human presenters able to convey flawless human-like voice speeches for different use cases (digital marketing, corporate communications, employee training, etc.). On the other hand, "deep fake" videos can fool even expert personnel. Fortunately, plenty of software tools are available for digital image forensics. These tools support journalists to detect if a photo or a video has been manipulated or faked. This is particularly important in case of user-generated

---

[451] Kenneth E. Foote, "To Remember and Forget: Archives, Memory, and Culture", The American Archivist, Vol. 53, No. 3 (Summer, 1990), pp. 378-392 (15 pages), Published By: Society of American Archivists

[452] Reuters Staff, "Reuters applies AI technology to 100 years of archive video to enable faster discovery, supported by Google DNI", AUGUST 13, 2020, Reuters web page: https://www.reuters.com/article/rpb-lavita-video-archive-idUSKCN2591VO

[453] Lily Kuo, "World's first AI news anchor unveiled in China", 9th November 2018, The Guardian, https://www.theguardian.com/world/2018/nov/09/worlds-first-ai-news-anchor-unveiled-in-china

content (UGC). Furthermore, AI-based technologies can solve several problems: denoising, super-resolution, etc.[454].

Research challenges

News production is a quite wide process involving several tasks and issues. By focusing on the production of audiovisual news (e.g. TV news programs), research challenges have still to be addressed in the following areas:

Video compression. Video standards have existed since the early 1990s. The demand for multimedia has increased and a huge amount of data is being produced. So, efficient communications need to advance compression technology. Neural network "encoders" are challenging the conventional approach. AI-based compression can provide the same level of visual quality with fewer bits[455,456,457,458,459,460].

Information retrieval. It is a hard task for computer systems to analyze, understand and represent contents in natural language. Neural networks have been applied to this task and need more investigation and trial in the professional media domain. AI and ML technologies (in particular "deep learning" technologies) can support different tasks:

- ***Natural language processing***. This allows to better understand and model the user information need by exploiting the huge amount of corpora of information.
- ***Image/video recognition***. This allows to extract features and search a multimedia corpus of information (e.g. objects and entities involved rather than just pixel and colour-related information).
- ***Knowledge representation***. This allows to build better data structures and search algorithms to identify meaning, synonyms, and relations between terms and concepts.
- ***Learning***. This allows to learn relevance ranking functions, to classify query intent and documents and to offer personalised results.

Automatic Speech Recognition (ASR), also known as Speech to Text. AI-powered speech recognition technology has the power to convert speech content to text and to recognise an individual based on their voice command.

---

[454] U. Schmidt and R. Stefan. "Shrinkage Fields for Effective Image Restoration". IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
[455] D. Ding, Z. Ma, Di Chen, Q. Chen, Z. Liu, and F. Zhu. "Advances in Video Compression System Using Deep Neural Network: A Review and Case Studies". January 2021, arXiv:2101.06341v1
[456] E. Çetinkaya, H. Amirpour, C. Timmerer, and M. Ghanbari, "FaME-ML: Fast Multirate Encoding for HTTP Adaptive Streaming Using Machine Learning", IEEE International Conference on Visual Communications and Image Processing (VCIP), 2020.
[457] Cisco Annual Internet Report, 2018–2023
[458] M. Nagel, M. van Baalen, T. Blankevoort, M. Welling. "Data-Free Quantization through Weight Equalisation and Bias Correction." IEEE International Conference on Computer Vision (ICCV), 2019.
[459] A. Kuzmin, M. Nagel, S. Pitre, S. Pendyam, T.Blankevoort, M. Welling. "Taxonomy and Evaluation of Structured Compression of Convolutional Neural Networks.", 2019, CoRR, arXiv:1912.09802
[460] QUALCOMM, "How AI research is enabling next-gen codecs" (2021): https://www.qualcomm.com/news/onq/2021/07/14/how-ai-research-enabling-next-gen-codecs

Generative Adversarial Networks (GANs). Deep learning technologies can synthesise highly convincing images and voices. However, natural video generation is still lagging.

Societal and media industry drivers

Vignette 1: AI-enabled live news coverage of a flood event in Sarno, Italy

Sarno (Italy). It's been raining for around 72 hours. Maria Luisa works as a journalist for RAI, the Italian national broadcaster. Today, due to some inconveniences along the road, she decided to work from home. At 3:14 pm (local time), while she is in front of her PC, a roar suddenly startles her. Looking out of the window she is completely shocked. It's raining heavily now and the roar's cause is evident: a landslide has broken off the hills right in front of her house. Immediately, Maria Luisa warns the police about the occurred event and a few seconds later she alerts her newsroom. To save time, Maria Luisa uses her mobile phone for video calling her colleagues and shows the landslide. Likely due to the bad weather conditions, the performance of Internet connection is low. Nevertheless, AI-powered algorithms optimise bandwidth usage while maintaining an appropriate level of quality. So, Maria Luisa and colleagues of her, decide to go live on the "Rai News 24" channel. From the balcony of her house, Maria Luisa can broadcast “live” the severity of the event: it rains a lot and a landslide just broken off and is visible on the hills.

Within a few minutes, the RAI control room (central newsroom) is overwhelmed by a multitude of reports and videos shot from local citizens thus witnessing what is happening at Sarno and at surrounding areas. Supported by AI tools, RAI's newsroom can manage this huge amount of information and automate different tasks: metadata extraction, video enhancing, scene identification, etc. In this way, RAI can broadcast and update in real time a map of Sarno and surroundings showing the ongoing situation (streets and/or bridges closed, dangerous areas, etc.). Videos sent from citizens can be verified, selected and broadcasted live. So, citizens receive live fresh information about what is going on in their city.

Due to bad weather conditions, no drones or any other aircrafts can fly and take aerial images. Nevertheless, available Digital Surface Models (DSMs) of that area allow technicians within the newsroom to build a 3D visualisation. AI-powered tools allow newsroom staff to generate a virtual simulation of the landslide that occurred a few minutes ago. This simulation is live broadcasted thus enhancing the storytelling. This virtual representation highlights the actual risks associated to possible further landslides that might break off in the next hours. Live news broadcast contributes to speed up the reaction of citizens. Local authorities can easily begin to evacuate several areas. At 6:04 pm (local time), a strong roar occurs: a second big landslide breaks off from the Pizzo D'Alvano mountain nearby. Two million cubic metres of mud fall upon Sarno. Fortunately, most houses had been evacuated and many human lives have been saved.

The day after it is not raining anymore. By searching in internal archives, RAI journalists do not find any similar event to have occurred in that area. The only aerial footage of Sarno is that recorded from a helicopter for the Giro d'Italia (cycling race) some years before. These footages document the urban and environmental conditions before the disaster. AI technologies can analyse these videos and extract relevant information (e.g., flight path, altitude, speed, camera angle, etc.). This information can be exploited for managing flight and shooting parameters of a drone that can document the current situation. By reproducing compatible manoeuvres of the helicopter, the drone's footage can highlight the damages caused from the event that just occurred. The day after, Maria Luisa shows the area covered by the mud. She can investigate the reasons and the responsibilities why the event has occurred.

Future trends for the media sector

Despite the concerns regarding the harmful use of these technologies (e.g. "deep fakes"), AI-powered tools will enhance the production of news. Developed for solving repetitive tasks, these technologies will be more and more used for freeing up time for novel "creative" skills.

By enhancing video compression tasks, AI technologies will enhance the streaming of videos on mobile devices. Real-time and high-quality video news will be more and more watched on small and portable devices. These growing consumer expectations will be met by the future availability of 5G technologies that will increase the speed of the wireless networks.

Newsrooms will elaborate on more and more amounts of information, related to facts that occurred locally and worldwide. Journalists will rely on AI-powered tools for news reporting and/or investigative projects[461]. AI technologies will be used for automatic content indexing and retrieval, detection (e.g., people, places, objects, concept, etc.), text/video summaries generation, video editing, generation of synthesised video (e.g., virtual presenters), translation in different languages. For a media company, only contents really matter. So, news will be delivered on any existing platform. Even on a still not fully clear one, like the Metaverse.

Goals for next 10 or 20 years

AI technologies have shown to evolve so rapidly that it is difficult to predict what will happen over the next 10-20 years. Newsrooms will rely on AI technologies for enhancing and changing internal workflows. Most repetitive tasks (video editing, summary generation, translation and delivery in different languages, etc.) will be performed by software tools. This opportunity will allow journalists to focus mainly on ongoing investigation/projects. To better meet consumer expectations, newsrooms will establish new synergies with citizens. AI-powered technologies will support information (and video) exchange guaranteeing ethical and professional aspects. To enhance internal training, newsrooms will foster new partnerships and joint projects with scientists and researchers. By the usage of advanced technologies, newsroom personnel will perform more investigations thus returning a deeper information to final users.

[461] S. McGoey, "A decade of digital evolution to help reporting revolutions at ICIJ", May 25th 2021, ICIJ website, https://www.icij.org/investigations/panama-papers/a-decade-of-digital-evolution-to-help-reporting-revolutions-at-icij/

9.3 Robot journalism

***Contributors**: Mike Matton (VRT)*

Current status

Newsrooms are heavily under pressure due to several reasons. First of all, the amount of news and information generated worldwide is exploding. Journalists have to be selective on the information they process. The rise of social media and other platforms has also contributed to this information explosion of the past decade. Secondly, news needs to be produced and distributed faster than ever. The competition to be the first to publish a news item is higher than ever before and sometimes depends on seconds. Finally, the pressure to produce more content is always there. In order to retain consumers, the offering needs to be large and pretty complete.

Over the past few years, advances in AI have caused AI technologies to make their way into the newsrooms. In the narrow sense of the word, "robot journalism" a.k.a. "algorithmic journalism" refers to stories automatically generated by AI algorithms, typically derived from one or several data sets. Lewis et al. define it as "*any process or system of news production under the control of media or electronic devices, with little or no external influence*"[462]. Several organisations worldwide have been using robot journalism to produce part of their content offering. Examples include: the Associated Press, which uses the Wordsmith platform to write compelling narratives on financial recaps; PA Media[463], which developed the RADAR platform to generate data driven news articles on sports, financial data and elections; Forbes, LA Times, etc. Robot journalism is not only limited to large media organisations; smaller organisations such as MittMedia in Sweden[464] have also been using it in practice for the automatic generation of Real Estate articles.

Initially, robot journalism was limited to extracting data from databases and using this data to automatically fill in the spaces in pre-defined template articles. Over the course of the past years however, those technologies have become smarter, with deeper data analysis taking place and the creation of more complex narratives in a story. However, applications are still mostly limited to domains with a lot of structured data available such as sports or financial news. There are currently promising systems available that can automatically generate narrative summaries of sports games. Producing more complex stories or narratives is however still a challenge. Although there are some nice examples[465] around, the creation of those stories remains in the hands of journalists.

---

[462] Seth C. Lewis, Andrea L. Guzman, and Thomas R. Schmidt. 2019. Automation, Journalism, and Human–Machine Communication: Rethinking Roles and Relationships of Humans and Machines in News. Digital Journalism7, 4 (April 2019), 409–427.

[463] Radar- Combining the latest in AI with skilled writers to dynamically create high-quality content at massive scale: https://pa.media/radar/

[464] L. Southern, Robot writers drove 1,000 paying subscribers for Swedish publisher MittMedia (2019): https://digiday.com/media/robot-writers-drove-1000-paying-subscribers-swedish-publisher-mittmedia/

[465] Guardian, A robot wrote this entire article. Are you scared yet, human? (2020): https://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3

The main idea behind robot journalism is that through the automated creation of articles, journalists in the newsroom will have more time to invest in developing more complex news stories, currently out of scope for automated journalism systems. In this way, some of the challenges mentioned earlier can be tackled. In practice however, several newsrooms have fired journalists because their work was now done by robot journalists. The most known example is Microsoft replacing 27 journalists by an AI system in 2020[466].

In newsrooms, "robot journalism" usually has a broader interpretation. It not only stands for completely automated news articles but also for AI and journalist working together in hybrid ways, where AI systems help journalists to navigate information, write parts of articles, assess veracity of information, create summaries etc. In this broader definition of the term, many more newsrooms worldwide are planning to or in the process of integrating such technologies into their workflows.

Research challenges

Most – if not all – of the robot journalism systems heavily rely on the availability of structured data. Unless this structured data is available, the systems will fail or produce unsatisfactory results. Also, the transformation of unstructured data into useful semantic structured data is not a simple task. Fundamental as well as applied research into technologies capable of doing this remains necessary. It requires analytical processing of unstructured information. Humans are much better at analytical thinking and reasoning with unstructured data compared to computers.

In news production, the elements of surprise and creativity are also important. As of today, human journalists are much better in creative thinking and surprising the reader compared to computers. Although advances in artificial creativity are being made, they still are not convincing enough to replace journalistic work.

Furthermore, there is the language challenge. As robot journalism heavily relies on natural language processing/generation systems, their performance is also language dependent. Those systems work much better in major languages such as English compared to smaller languages.

Finally, assessing the veracity of information, and making conclusions on whether or not to publish a story is still a challenge in itself. Disinformation is in fact one of the major hurdles to overcome. AI systems that automatically generate trustworthy news can often be misused to generate disinformation. This is in fact already happening in some fake news farms. Furthermore, tampering with the data sources themselves can remain undetected and generate fake articles without anyone noticing before it is in place. There is still a lot of work on rigid editorial processes and trustworthiness assessment of information before such systems can be used at mass scale.

466 J. Waterson, Microsoft sacks journalists to replace them with robots (2020): https://www.theguardian.com/technology/2020/may/30/microsoft-sacks-journalists-to-replace-them-with-robots

A good overview of current research challenges can also be found in more detail in a recent article by Kotenidis et al[467].

Societal and media industry drivers

Vignette: Robot journalism for Olympic Games reporting

Chantal is editor in chief of a large newspaper. Thanks to the digitisation of the past two decades, audiences are more and more reached online with respect to the traditional newspapers. This also brings challenges, however. In the digital world, more content is needed compared to the regular newspaper, and people want to get updates regularly. The traditional collection of articles in a typical newspaper has been replaced by a continuous feed of articles. Readers browse and scroll for interesting content throughout the day, similar to how social media feeds work.

There is a major challenge for Chantal ahead, the Olympic Games. Her newspaper has always been the reference for Olympic Games reporting and she wants to keep this reputation. She has a decent team of journalists and reporters, who will work both on premise as well as in the newsroom. Although her team has a decent size, the team is unable to report on every discipline and every game. Furthermore, reporting is not limited to the Olympic Games period itself. Even in the months before, a lot of stories and previews need to be published in order to raise the attention of the audience and raise the excitement. Even though her team is working at maximum capacity, there is too much to report on, and Chantal knows some readers will be disappointed if their favourite sports is not reported and as a result might move away to other media channels. Moreover, the audience follows the newspaper not only through their website, but also, more importantly, through various social media channels. This creates an additional stress on the team, as they now not only have to cater for the newspaper, their own news website, but also keep the social media channels active and filled with interesting content.

Luckily, the reporters are not the only resource Chantal has available. Her newspaper has bought access to several data feeds on the Olympic Games, with statistics about the athletes, loads of historical information, and live data feeds with events from every single competition in the Olympic Games. Publishing this data as such is not compelling to the audience. Therefore, the newspaper bought a license for a robot journalism system. One year ago, when Chantal was first introduced to the system, she was impressed with its ability to produce live reports on sports games in many different formats and publish them on a variety of channels, including social media feeds. With the right configuration in place, hundreds of short reports, live game reports and highlight overviews are generated automatically. Thanks to the powerful linguistic engine of the system, the articles read smoothly and fluently. Chantal is keen on the fact that the decision to publish is still in the hands of her editorial team. However, thanks to the powerful robot journalism system, these decisions are made efficiently, allowing her and her team of reporters and journalists to focus more on the major highlights of the event.

[467] Kotenidis, Efthimis, and Andreas Veglis. 2021. Algorithmic Journalism—Current Applications and Future Perspectives. Journalism and Media 2: 244–257. https://doi.org/10.3390/journalmedia2020014

Although Chantal is happy with the current situation, which keeps her newspaper on the competitive edge, she knows even more will be necessary to keep her newspaper relevant and to retain its consumer base. She hopes that in the near future AI systems will be able to offload more work of the editors such that they can focus on the real cutting edge quality content.

Future trends for the media sector

We foresee that in the near future, every newsroom worldwide will make use of AI technologies in one or another way. Advances in AI technologies will almost certainly help newsrooms to be more efficient in their work, with repetitive and less complex jobs fully automated as discussed before.

However, the replacement of a human journalist by a computer is still a long way ahead. It is even doubtful this will ever happen. The main trend we foresee for the future is the hybrid collaboration between journalist and AI system. The places where AI technologies can augment the work in the newsroom are numerous. Figure 87 presents a (non-exhaustive) overview of processes in the newsroom that can be supported by AI technologies. There is potential throughout the workflows, from ideation over research, production, publication, feedback and interaction, and archival.

For this purpose however, journalists need to adapt their way of working, and learn how to rely on the available AI systems in the newsroom and beyond. More specifically, they need to understand the strengths and limitations of the AI systems. A huge challenge ahead in this is not only the integration of the technology into the newsroom, but also the whole process of change management to adhere to new ways of working. If working well however, these trends will make the newsroom much more efficient, both in terms of quantity as well as in quality of the news stories provided.

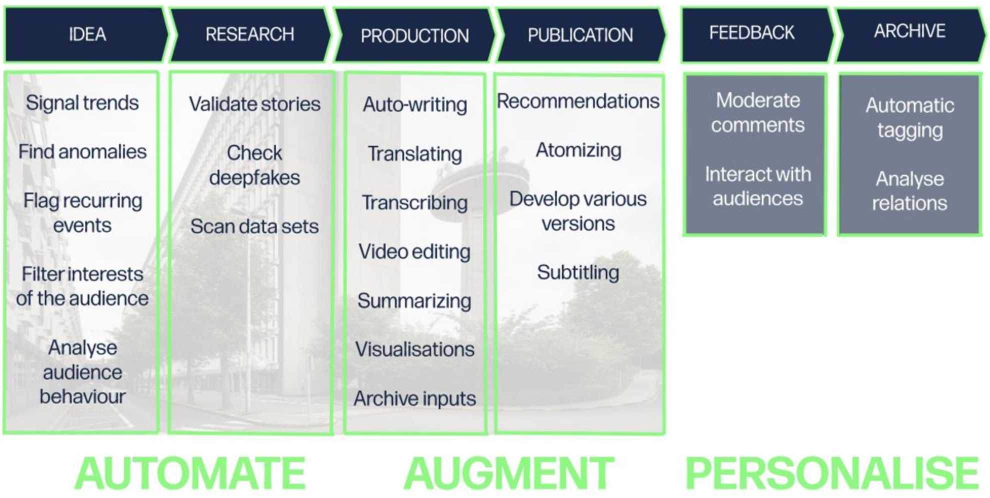


Figure 87: AI assisted newsrooms.

Furthermore, the consumption of news is also diversifying and becoming more personalised. For this purpose, newsrooms need to publish stories in many different formats and on many different platforms. Also for this purpose, AI technologies can certainly assist to generate the many different versions of the content to publish.

Goals for next 10 or 20 years

In the next 10 or 20 years, robot journalism will gradually be introduced in newsrooms. It is not expected however that robot journalism will completely replace journalists due to the complex nature of news production, information assessment and reasoning. Journalists will gradually adapt to work together with AI systems that help them in their daily job.

The successful newsroom of the future will seamlessly integrate fully automated robot journalism and hybrid (AI-assisted) journalism, able to produce a rich content offering on many different platforms, and personalised to the needs of the news consumer.

9.4 AI for counteracting disinformation

***Contributors**: Deutsche Welle (DW)*

Current status

The phenomenon of online ***disinformation*** has evolved since around 2010 and is defined as "*false, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit*"[468]. While the spreading of false or manipulative information has occurred for centuries, the significance and negative impact of this activity/phenomenon has increased with the emergence of social media, digital information production and consumption as well as advances in technology, including Artificial Intelligence. Although the effects of online *disinformation* have been addressed by fact checking and verification specialists for almost one decade, events such as the US presidential election in 2016 and the Covid-19 pandemic have brought the significant risks for society, democracy, and individuals to mainstream, academic and political attention.

Many different stakeholders are engaged in counteracting *disinformation:* not only social media platforms, fact-checking initiatives, open-source intelligence specialists and news media organisations, but also academia, governments, educational institutions and civil society initiatives. One or more of the following interrelated approaches come generally into use for the purpose of counteracting *disinformation*:

[468] There are many definitions for *disinformation.* We have chosen the one first defined by Wardle, Claire, and Hossein Derakhshan. "INFORMATION DISORDER: Toward an interdisciplinary framework for research and policy making." (2017). This definition is also used in the report issued by the European Commission's High Level Expert Group on Fake News and Online Disinformation (Source: European Commission. "A Multi-Dimensional Approach to Disinformation: Report of the Independent High-Level Group on Fake News and Online Disinformation." (2018).

- Verifying content (e.g., videos, photos, or posts) and social media accounts;
- Checking statements (claims) made by public figures against facts;
- Identifying disinformation narratives/stories in social media;
- Conducting media literacy and education/training programmes;
- Establishing self-regulation schemes and regulatory frameworks;
- Developing counteractive methods, technologies, and support tools.

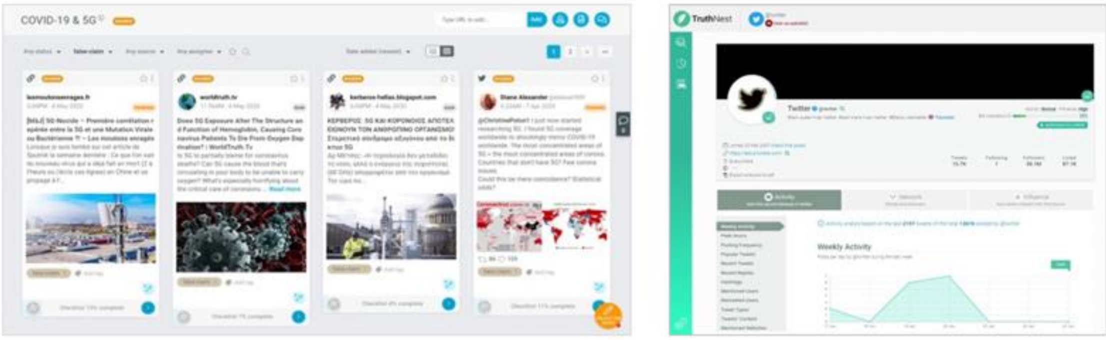


Figure 88: Examples for current support tools to counteract disinformation: Truly Media[469] (left) and TruthNest[470] (right).

For many years, AI technologies have played an important role in counteracting *disinformation*, especially in tools and systems used for content verification, fact-checking and social media disinformation analysis (Figure 88). The need for AI support has recently increased: On one hand the frequency and scope of disinformation has grown to a level that manual approaches cannot handle. On the other hand, adversaries use advanced AI technologies and automation for targeted campaigns, content manipulation or synthetic media production, which in many cases are only detectable with AI-powered systems.

Despite multiple AI based support functions being available, there are several shortcomings, limitations, and missing elements to ensure long-term success in counteracting online disinformation. The ***current areas of limitation*** are presented in Figure 89 and discussed below.

AI functions and solutions. Most AI solutions today are good at specific, narrowly defined tasks that can help to identify disinformation elements and claims in social media. Examples are reverse image and geo search, detection of bot accounts, comparing digital content for detecting changes/manipulation (e.g., text, video, and photos), detecting deepfake face-swaps in videos or photos, analysing audio to detect manipulation, scanning large data repositories for specific keywords, and analysing certain aspects of content in social networks and relationships between accounts. **What AI cannot yet deliver for practitioners in fact checking and verification is the detection and analysis of entire, complex disinformation narratives,**

[469] Truly Media: https://www.truly.media/
[470] TruthNest: https://www.truthnest.com/

handling more complex tasks across social/digital platforms and involving multimodal data types, and covering all aspects and types of synthetic media detection/manipulation.

Underlying datasets. Although there is research into multimodal approaches[471], at present, the underlying datasets for AI solutions that are *practically* used to counteract *disinformation* often relate to one data type (e.g., either text, video, images, or audio) or only one content source (e.g., Twitter). Further, it can be generally challenging to collect quality datasets[472] and in many cases, they are related to only one domain (e.g., politics). In addition, the usage of certain datasets is subject to ethical concerns and many datasets are difficult to obtain (and to maintain for longer periods of analysis) in the light of regulation, Intellectual Property Rights (IPR), ethical requirements and the terms and conditions from media platforms. **Despite regulatory initiatives, there are currently limitations regarding datasets that are easily and openly available to those researchers/developers that produce AI solutions against *disinformation*, but which are yet ethically and legally compliant. Further, there are requirements for multimodal, cross-platform and multi-domain datasets.**

Human-AI Collaboration. Current AI-powered tools largely provide machine support for humans who need to conduct complex fact-checking and verification workflows. They enable otherwise (humanly) impossible analysis (detection) and reduce time, stress, or cost, but are largely based on manual human oversight and manual pre-detection (e.g., presenting the AI-function with a video in which a deepfake face-swap is suspected). Where semi or full automation could be technically achieved, there is usually associated human distrust in the capability of the AI-powered solutions to make an accurate and/or contextually acceptable decision. **So far, there are limited approaches for true human-machine collaboration or (socially acceptable) forms of automation.**

Responsible and Trustworthy AI. Most current AI functions and services used in the context of counteracting *disinformation* are accuracy and performance oriented, with little or no information given (by third-party providers) about the AI function/model itself, its legal compliance or measures taken to provide explainability, to mitigate bias or to increase reliability/robustness. Such limitations related to *Responsible/Trustworthy AI* can impact on successfully counteracting *disinformation* for various reasons: 1) In this domain, many decisions are related to the comparatively vague and complex concept of "truth". 2) Unlike some commercial AI application domains, the work of fact checkers, verification specialists or journalists is also influenced by immaterial aspects (e.g., societal and public value systems). 3) It is typical for fact checkers, verification specialists, journalists, or open-source intelligence analysts to be curious, show attention to detail and question any presented information prior to further using it. 4) All stakeholders (specialist staff, editorial managers or the board of management of the organisation) are bound by editorial control rules (e.g., dual control principles), journalistic codes and specific organisational values as well as legal frameworks

[471] An example for such research is: A. Giachanou, G. Zhang, and P. Rosso. "Multimodal multi-image fake news detection." 2020 IEEE 7th International Conference on Data Science and Advanced Analytics (DSAA). IEEE, 2020.
[472] F. Torabi Asr, and M. Taboada."Big data and quality data for fake news and misinformation detection." *Big Data & Society* 6.1 (2019): 2053951719843310.

related to publishing/journalism. **Current shortcomings related to the integration of Trustworthy AI principles into AI functions that help to counteract *disinformation* can raise (justified) questions among affected stakeholders and therefore reduce their acceptance and use by related specialists and/or their organisations.**

Usability and User Experience. While there are (and will be) many useful stand-alone AI functions or solutions available, it remains difficult to transform their technical output/predictions into suitable user interfaces within the tools used by practitioners and to create satisfactory user experiences that are adequate for non-technical users in counteracting *disinformation*. **This difficulty is due to a gap of knowledge and funding resources that occurs between any existing (or future) AI function presented in the format of code, a dockerised container or Application Programming Interface (API) and the user interface of an end-user tool that makes the result of this function usable.**

End user tools and systems. Various specialist initiatives and projects conducted research and delivered AI-functions for counteracting *disinformation* at the level of research outcomes, piloted prototypes, or open-source solutions. The (global) supply market is highly fragmented and consists of many small (or even ultra-small) players. There are few European commercial solutions in the market specifically targeted at the fact checking and verification workflow (e.g., such as Truly Media[473]). Other off-the-shelf products in that direction are headquartered outside Europe, are not yet commercially mature enough or target a different customer base (e.g., corporate reputational analysis and PR). **There remains a lack of tailored, end-to-end, and mature products for all relevant stakeholders that are suitable (in business terms), accessible (in financial and integration terms) and sustainable (in both maintenance and energy terms).**

Figure 89: Areas of current limitation for counteracting disinformation with AI.

Research challenges

While existing AI approaches and tools are already invaluable to counter *disinformation*, there is a need for improvement in areas such as AI technology advancement, datasets, human-AI collaboration, trustworthiness, user interface transfer and industry products. These research

473 Truly Media: https://www.truly.media/

areas are presented in Figure 90 and discussed below. To develop AI systems for countering disinformation that have this wider capacity, further research is needed over the next decade to overcome limitations. This research is required from diverse academic fields, related to technology, business, and society. Due to the fact that this chapter focuses on AI technologies, other important research challenges are not covered, e.g., cultural, psychological or cognitive aspects. The following research areas and challenges should be addressed:

AI Technology Advancement: This research area relates to next generation AI approaches and functions that fill current technology gaps in counteracting *disinformation*. Examples for research subjects are:

- Multimodal content analysis (e.g., image with integrated text)
- Cross-platform content and network analysis
- Linguistic and country-specific environment analysis
- Detection of content manipulation by means of synthetic media
- Automatic synthetic content detection/flagging
- Dynamic AI-updates in counteraction tools (to match disinformation actors)
- Early detection of arising disinformation narratives/elements
- Causal, contextual, and cultural analysis of complex statements
- Analysis of complex narratives / disinformation stories over time
- Automatic identification of check-worthy, potentially harmful elements
- Integrated analysis with Blockchain based authentication approaches.

Next Generation Datasets: This research area relates to next generation datasets used for the training and evaluation of AI functions that help to counteract *disinformation*. It involves technology research (e.g., synthetic data) and societal research (e.g., policies enabling long-term access to social media platform data). Examples for research subjects include:

- Multimodal and multilingual datasets
- Cross platform datasets
- Datasets that enable early or even real-time detection
- Synthetic datasets (overcoming issues of real datasets)
- Legal, ethical and IPR compliance certification for datasets
- Regulated datasets for specific uses/users (public value)
- Specialised datasets for *disinformation* detection purposes.

Human-AI Collaboration and Automation: This research area relates to enabling true human-AI collaboration and acceptable automation of fact checking and verification workflows in terms of journalistic/content environments, legal, ethical, and business issues. The research involves the fields of AI technology, human-computer interaction, interface design, AI-based product design and trustworthy AI. Examples for research subjects include:

- Automatic filters to select suspicious content
- Automation & collaboration approaches to fact checking/verification

- Role of humans / human-in-the-loop / oversight
- Workflows with no, minimal, semi or full automation
- Seamless conceptual integration of the above

- Resolution of editorial/legal responsibility conflicts (human vs machine)
- Issues of censorship and freedom of speech related to the use of AI
- Issues of editorial control, journalistic values, and legal frameworks
- Role of Trustworthy AI in Human-AI collaboration in *disinformation* domain
- Characteristics of "acceptable" automation in the *disinformation* domain.

Trustworthy AI Capability: This research area relates to increasing the overall transparency of AI functions used in the context of counteracting *disinformation* and integrating specific trustworthy AI approaches/tools to enable a responsible and accepted use of AI in this field (and better human-AI collaboration, see point 3 above). Examples for research subjects include:

- Role of transparent/trustworthy AI in the acceptance of AI tool support
- Tailored AI transparency certifications (provider, model, data, legal)
- Tailored Trustworthy AI certifications (explainability, fairness, robustness)
- Translation of trustworthy AI output for interfaces / non-technical users
- Balancing decisions between effectiveness and trustworthiness
- Can transparency/trustworthy elements enable full automation?
- How to avoid misuse of AI technologies employed against *disinformation.*

Function-to-Interface Transfer: This research area relates to bridging the knowledge gap between a delivered AI function and the user interface of an end-user tool that makes the result of this function well usable for non-technical users. Examples for research subjects include:

- Expertise and staff roles to overcome this challenge (user-side)
- Alternative ways of presenting output of AI functions (provider-side)
- Translation of AI output for interfaces and into user language
- Translation of trustworthy AI output for interfaces / non-technical users
- Personalised approaches: matching AI affinity/expertise of end users
- Dashboard approaches for AI analysis outcomes
- Reduction of complex AI analysis outcomes.

Tailored European Products: This research area relates to the market for AI-powered tools used in fact checking and verification workflows and ways of enabling dedicated, tailored European, end-to-end products for this purpose that are suitable and accessible for large and small European professional stakeholders (e.g., fact checking organisations, media companies, self-employed journalists, or civic initiatives). Examples for research subjects include:

- Existing AI-powered tools/functions and their providers
- Multilingual products
- Product characteristics for realistic adoption by users/organisations
- Opportunities/barriers: European public sector / public service products

- Opportunities/barriers: European commercial products
- Multi-faceted products: one AI-powered back end with multiple front ends.

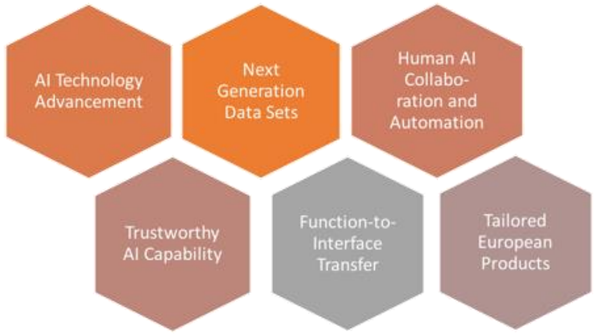


Figure 90: Suggested research areas to support the counteracting of disinformation with AI.

The scenario below illustrates a vision for counteracting *disinformation* in 15-20 years from now, from the perspective of a media industry stakeholder and related to the six research areas suggested above. To realise this vision, it is important to achieve improvements in the aforementioned areas. The scenario also indicates the overall societal and media impact that the successful realisation of such research activities could have.

Societal and media industry drivers

Vignette: Counteracting disinformation in the 2030s with the AI-powered CADI-Tool

Carmen is a freelance medical journalist, working in the media environment of the late 2030s. During her 20-year career in this sector, she has seen continued growth of online *disinformation*, affecting all genres of published content on any platform. Not surprisingly, this led to major public upskilling programmes for information workers, journalists, pupils, and the public as well as regulatory measures and global agreements between media platforms and governments. At the same time, Carmen saw the development and uptake of complex AI-powered support systems for dealing with and counteracting *disinformation* daily.

For over 10 years now, Carmen has had a single-user subscription to a web-based product called *CADI*, which provides similar features as other commercial systems available that also integrate with corporate content systems. All these AI-powered systems for counteracting *disinformation* have in common that they are widely adopted and socially accepted, make use of synthetic or trust-certified datasets, automatically update to state-of-the-art functions (also to keep up with *disinformation* adversaries), come with transparency and trustworthy AI certifications tailored to this domain, provide easy-to-grasp assistance via personalised, visually dynamic and flexible end-user interfaces, based on a user experience that is driven by seamless human-AI collaboration, involving a high level of workflow automation and flexible levels of human oversight.

All types of information workers, including Carmen as a freelancer, can easily deal with all types of disinformation related workflows and tasks: from verifying content items, to checking claims against facts and analysing complex social media narratives, including those that are rapidly emerging.

It is Carmen's first task of the day to check the information agenda. She will then research, produce and submit by the end of the day a video story on emerging reports about a virus outbreak in a neighbouring country. Checking the news agenda is quickly done with her personalised *CADI* dashboard, already set to her preferences (mid-level information detail and low-level technology affinity). Carmen added two required languages and geographic regions, to achieve cultural and linguistic analysis matches, as well as the required content keywords related to her medical topic. Carmen quickly glances over the resulting data visualisations, showing in an integrated way the breaking news coverage, trending social stories around it, suspected disinformation narratives, already debunked claims and a list of key media items that are either shown as suspicious or already verified by other information workers. The latter list is divided into fully synthetic, synthetically manipulated, and non-synthetic media items. Based on the news (and disinformation) overview that she had obtained earlier via the dashboard, Carmen uses the *CADI* system to conduct a further universal search job across multiple media platforms and media types such as text, video, images, or audio. Apart from reporting the news of the virus outbreak, she will also contrast official statements with circulating theories and highlight selected *disinformation* elements as it is common practice. The *CADI* system acts as an early-warning system in this breaking news situation and automatically suggests "suspicious" statements, narratives, and media items for her (human) review and for use in fact-checking reports. Carmen is particularly pleased to have this function as it took over a decade for AI systems to identify what humans might regard as "suspicious". At the same time, to avoid overload, the system has automatically deleted several *disinformation* elements in her results feed, based on transparent, certified approaches she is fully aware of.

While Carmen is accepting some of *CADI*'s decisions related to *disinformation* elements (as she knows the AI-technology is trustworthy and has been certified), she decides to follow up the transparency and trustworthy AI information provided for others. One reason for checking some specific aspects manually is that the media editor to whom she submits her video report will do the same for the purpose of editorial control. In particular, she double checks the AI system's decision that a popular video featuring the health minister of the neighbouring country is a deepfake, because getting this wrong may have legal implications for the media company publishing her video.

After having spent a little too long on research, Carmen now quickly produces the video. Prior to finalising the video, she asks *CADI* for an update of both news and *disinformation* developments. Luckily, there weren't any major developments. She presses the submit button, leaves her desk and while she walks home, Carmen's mind wanders back to the early days of her career in the early 2020s. She can hardly believe that counteracting *disinformation* was a difficult, time-consuming and complex workflow, sometimes with limited success, not possible in live or breaking news situations, hindered by language barriers and conducted by a few specialists in the media sector, who played a game of catch-up with the ever-advancing disinformation actors.

Future trends for the media sector

Further technical advances will also help to drive the production and distribution of false or misleading information. However, conducting extensive multidisciplinary research into the next generation of AI-powered solutions as described above can lead to a turning point, giving way to a range of opportunities and benefits for the media industry – as well as other related domains that can benefit from similar functions/tools. The following future trends can be anticipated:

- Increasing acceptance of and trust in AI-powered tools for counteracting *disinformation* in media and society through ***transparency and trustworthy AI certifications***, that can be easily used by end-users, stakeholders, journalistic codes of conducts and regulatory frameworks.
- Removing barriers to workflow automation in an area that involves the complex concept of "truth" by establishing successful ***Human-AI collaboration models***.
- Providing significantly more information workers in media and society with access to powerful and suitable ***support tools*** to deal with and counteract *disinformation*.
- Enabling ***earlier or even real-time detection***, which solves the societal problem of reactive fact-checking and verification after *disinformation* has already spread.
- Allowing creative media and information workers to focus on their core tasks, while the complex, time-consuming ***workflows*** related to analysing *disinformation* are largely ***automated*** – in a responsible, trusted way.

Goals for next 10 or 20 years

By 2040, all relevant stakeholders will be involved in counteracting *disinformation*, ranging from information workers in all domains (media, government, business, and society), to members of the public and (global) media platforms – benefiting from widely automated, trusted and seamlessly integrated counteraction workflows, early identification of *disinformation*, and significantly more impactful, accessible, and user-friendly support tools.

By then, AI-powered support systems for counteracting *disinformation* will have capabilities such as

- Multimodal and cross-platform analysis;
- Linguistic, country, culture, and context analysis;
- Full synthetic content and synthetic manipulation analysis;
- Automatic and early (real-time) detection of *disinformation;*
- Automatic detection of check-worthy items, claims or narratives
- Seamless and flexible human-AI collaboration workflows;
- Certified information related to Transparency, Trustworthy AI, Datasets;
- Automatic technology upgrades to match tools of disinformation actors;
- Interoperability with content authentication systems (e.g., Blockchain-based).

The above capabilities are enabled on one hand by major advances in realising Trustworthy AI (accurate, performant technologies with yet trusted and explained outcomes) and on the other hand by widely available, ethically, and legally certified datasets that are needed for AI model training and evaluation for such functions. In combination, this is the basis for and can enable successful human-AI collaboration. Stand-alone AI technologies, functions and services will be integrated into tailored, user-friendly, and accessible support products targeted at information workers, which are widely available as off-the-shelf applications, affordable web-based subscription services or for seamless integration into corporate content management systems and their user interfaces. Specific public subsidy and co-funding programmes are in place to ensure access to these high-end, AI-powered systems for all types of users who need them. Core back-end technologies connect with multiple front ends for different user domains, where front ends are featuring high degrees of multi-faceted personalisation, dashboard views, fine-grained visualisation of AI-predictions and easy-to-grasp (and easy-to-accept) trust-related information, such as AI certifications or explanations for AI actions that can be understood by non-technical users).

On the way towards achieving these ultimate goals around the year 2040, the following milestone points can be defined as interim goals for the years 2025, 2030 and 2035 (Figure 91):

Milestone 5 years: By 2025, apart from continued advances in AI technology, support products and user experience design, the AI functions provided for the purpose of counteracting *disinformation* will be certified in terms of Trustworthy AI and based on tailored, ethically, and legally compliant (certified) datasets.

Milestone 10 years: The developments during the 2020s formed the necessary baseline for achieving more (acceptable) automation in fact-checking and verification workflows as well as true human-AI collaboration, which is largely in place by 2030. This achievement is also driven by the by then more powerful AI analysis capabilities, advances in datasets, and widely accepted, accessible support tool products with a strong focus on AI-results usability.

Milestone 15 years: By 2035, the progress described above now begins to show real impact, leading to a significant reduction of both *disinformation* itself and its negative effect on media and society. This is driven by a combination of factors: further technical advance and excellence of Trustworthy AI functions and underlying datasets in use, wide availability of user-friendly and accepted AI-support tools (including public subsidies) and the implementation of seamless human-AI collaboration, enabling largely automated workflows if and where chosen.

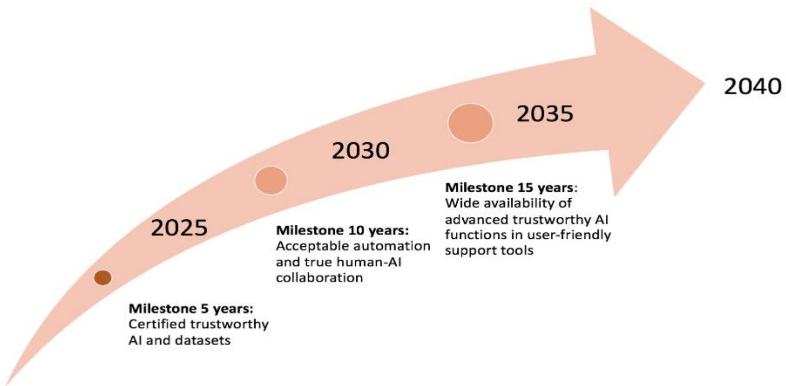


Figure 91: Milestones up to 2040 for counteracting disinformation with AI.

9.5 AI for next-gen social media

Contributors*: Symeon Papadopoulos (CERTH), Nikos Giatsoglou (CERTH)*

Current status

Social media, such as Facebook, YouTube, Twitter, and LinkedIn, have been around for at least 15 years, and have left an indelible mark on our online communication and socialisation patterns. As an example, chatting, voice messages, posts, and other means of communication are more popular today than phone calls, and users of all ages have been persuaded to let go of inhibitions and air their opinions confidently online. Owing to their popularity, social media have also fulfilled other roles besides communication. These include the role of information or education sources, entertainment in the form of games and funny videos, online marketplaces and bartering, job search and recruiting, citizen activism, connection with companies and institutions, and many more (Figure 92). Crucially, social media have had a huge impact on politics as well, as politicians routinely use them to shape their public image and exercise rhetoric online.

Despite their impressive applications, from a business perspective, the importance of social media lies in the availability of ***user data***, which offer rich and low-cost data on human behaviour. Such data can be mined with AI and ML techniques and monetised by social platforms to sell targeted ads to advertisers. If we had to pick a single application of ML to social networks, it would be ***recommendation algorithms*** to serve targeted ads, content, or friend recommendations to users. Although recommendation is not strictly necessary for communication, it is integral to social media as a practical way to explore the sea of user generated content in a personalised way. After all, it would be hard to imagine Facebook without its algorithmically curated “wall”. Recommendation algorithms are typically configured to increase the engagement and attention span of the users, encouraging them to spend more time in the social media, view ads, and make profit for the company. Unfortunately, this practice has also cast shadows over social media, as it can promote sensational content and leads to the creation of filter bubbles[474] and rabbit holes[475].

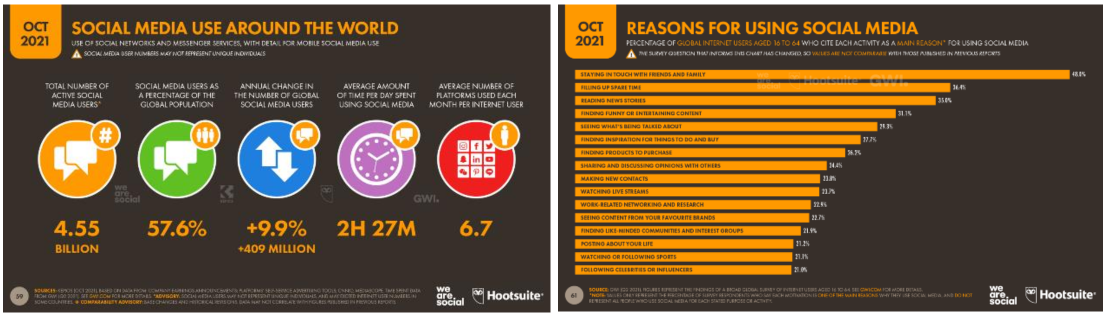


Figure 92: a) Social media use around the world in Oct. 2021 and b) main reasons for using social media[476].

From a technical perspective, recommendation algorithms[477] have made use of and fuelled interest in information retrieval, collaborative filtering, and graph processing. Information retrieval concerns the retrieval of documents from a document collection that is most relevant to a user query. In the case of recommendation, the query is the online profile and interests of a user, and the documents are the candidate contents. The objective can include other metrics beyond relevance such as novelty, diversity, and fair representation of content producers. In addition, collaborative filtering derives recommendations from users with similar interests, while graph processing factors the preferences of a user's friends. Link prediction is a specific problem where graph processing has had great success[478]. Beyond recommendation, other ML algorithms that had a big impact on social media draw from NLP, audio processing, and computer vision, and include:

- face recognition and automatic detection of people in images and videos;
- automatic cropping and captioning of images;
- speech-to-text conversion for video transcribing;
- automatic translation to conveniently access post in foreign languages;
- NLP and affective computing techniques for conversational agents, e.g., customer service bots.

Besides recommendation, social data has been an important and cheap source of social analytics and it was first used by businesses to gain feedback on brand perception and guide their

[474] Pariser, Eli. *The filter bubble: How the new personalized web is changing what we read and how we think*. Penguin, 2011.
[475] O'Callaghan, Derek, et al. "Down the (white) rabbit hole: The extreme right and online recommender systems." *Social Science Computer Review* 33.4 (2015): 459-478.
[476] Image taken from: DataReportal (2021), "Digital 2021 Global Digital Overview," retrieved from https://datareportal.com/global-digital-overview
[477] Kulkarni, Saurabh, and Sunil F. Rodd. "Context Aware Recommendation Systems: A review of the state of the art techniques." *Computer Science Review* 37 (2020): 100255.
[478] Wang, Peng, et al. "Link prediction in social networks: the-state-of-the-art." *Science China Information Sciences* 58.1 (2015): 1-38.

marketing strategies. Monitoring the social media is called *social media monitoring* or *listening*, and allows the users of related software to monitor keywords and topics across social platforms, time periods, and demographics[479]. These techniques have also been applied in politics, to gauge the public opinion and guide political campaigns, but with highly controversial results. Specifically, the Cambridge Analytica scandal has shown how user data can be exploited to influence major political events, such as the 2016 US elections and Brexit[480], raising serious ethical concerns on user data collection and causing backlash against social media. Despite widespread condemnation over the scandal, technology is now firmly intertwined with politics, with political parties routinely using bots for political campaigns and a shady industry developing around disinformation.

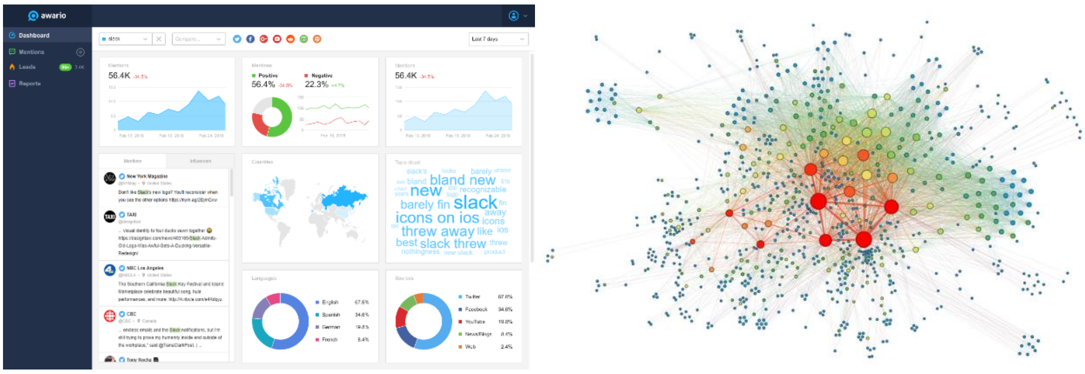


Figure 93: a) A social media analysis tool[481] (left) and social network graph visualisation[482] (right).

Despite the above misgivings, many applications of social media analysis of clear societal value have been proposed in the literature. These include:

- **opinion mining**: for capturing the public sentiment, emotion, or mood on contemporary topics in place of expensive user polling;[483]
- **smart city monitoring:** for detecting traffic jams, events, news stories, and trends;
- **environmental monitoring**: for monitoring the interaction of humans and nature, preserving biodiversity and tracking illegal trafficking;[484]

[479] Stavrakantonakis, Ioannis, et al. "An approach for evaluation of social media monitoring tools." *Common Value Management* 52.1 (2012): 52-64.
[480] Bömelburg, Raphael, and Oliver Gassmann. "Cambridge Analytica: Magical Rise, Disastrous Fall." *Connected Business*. Springer, Cham, 2021. 387-396.
[481] Image source: Social Media Today - https://www.socialmediatoday.com/news/12-of-the-best-social-media-analytics-tools-and-how-they-can-help-your-bus/546568/
[482] Image source: Wikimedia - https://commons.wikimedia.org/wiki/File:SocialNetworkAnalysis.png
[483] Cortis, Keith, and Brian Davis. "Over a decade of social opinion mining: a systematic review." *Artificial intelligence review* 54.7 (2021): 4873-4965.
[484] Toivonen, Tuuli, et al. "Social media data for conservation science: A methodological overview." *Biological Conservation* 233 (2019): 298-315.

- **cultural heritage monitoring:** promoting preservation and sustainable urban development;[485]
- **mental health monitoring:** for the detection of anxiety, depression, and suicidal tendencies[486], monitoring for adverse drug reactions;[487]
- **disaster management:** for propagating information during / after a catastrophe, coordinating recovery procedures.[488]

These applications use advanced NLP techniques and computer vision to extract information from unstructured data found in social media. The information from multiple domains is then fused and processed with appropriate ML algorithms, such as neural networks, in order to extract the desired output. As was the case with the business and political applications, collecting data without compromising the user's privacy is a key challenge.

In addition to data mining, two important trends for social networks have been emerging which may play a role in future media analysis: i) decentralised online social networks (DOSNs)[489] and ii) the *metaverse*[490].

Regarding DOSNs, although they were initially conceived in the heyday of p2p networks in the early 2000s, they have seen a surge of interest due to the concerns raised by centralised social media and the popularity of blockchain. DOSNs intend to operate without central servers and give back control to the users over their own data. This implies letting users explicitly grant consent to applications over their data (which can be revoked at any time) and/or monetising them, often through cryptocurrencies. Key actions and proposals towards the decentralisation of the web include the 2018 decentralised web summit[491], Tim Berner Lee's Solid project[492], and Elastos[493]. A key consequence of DOSNs is that they will place strong emphasis on decentralised and federated learning techniques, which are also gathering research interest lately.

Regarding the metaverse, it refers to a virtual world running parallel to the physical one, which we will access through our avatars or digital twins. At this moment, the concept is not precise but it has been fuelled by the popularity of VR and AR technologies and the recent commitment of Facebook to implement it, after the company was renamed to Meta. The metaverse promises

---

485 Liang, Xiaoxu, Yanjun Lu, and John Martin. "A review of the role of social media for the cultural heritage sustainability." *Sustainability* 13.3 (2021): 1055
486 Chancellor, Stevie, and Munmun De Choudhury. "Methods in predictive techniques for mental health status on social media: a critical review." *NPJ digital medicine* 3.1 (2020): 1-11
487 Sarker, Abeed, et al. "Utilizing social media data for pharmacovigilance: a review." *Journal of biomedical informatics* 54 (2015): 202-212
488 Saroj, Anita, and Sukomal Pal. "Use of social media in crisis management: A survey." *International Journal of Disaster Risk Reduction* 48 (2020): 101584
489 Masinde, Newton, and Kalman Graffi. "Peer-to-Peer-Based Social Networks: A Comprehensive Survey." *SN Computer Science* 1.5 (2020): 1-51
490 Lee, Lik-Hang, et al. "All one needs to know about metaverse: A complete survey on technological singularity, virtual ecosystem, and research agenda." *arXiv preprint arXiv:2110.05352* (2021)
491 "Decentralized Web Summit 2018," https://decentralizedweb.net/, 2018
492 Sambra, Andrei Vlad, et al. "Solid: a platform for decentralized social applications based on linked data." *MIT CSAIL & Qatar Computing Research Institute, Tech. Rep.* (2016)
493 Elastos: https://www.elastos.org/

to be an evolution of social media and a much more immersive experience, where people will interact through their avatars, generate VR content, attend events, and engage in online purchases with bitcoins. If this vision succeeds, it will offer unprecedented opportunities for data analysis, require even more invasive data collection as well as new AI algorithms for the operation of digital twins and the management of Non Playable Characters (NPC). On the other hand, critics point out the similarities with older efforts such as SecondLife[494] (Figure 94) and the timing of the Facebook announcement when the company is facing antitrust allegations[495] and younger populations are migrating to other platforms[496]. Whatever the case, the importance of VR and holographic content in social media will undoubtedly increase.

Figure 94: A social gathering on the metaverse Second Life [497].

Research challenges

While social media analysis presents many opportunities, it also presents important challenges. These include the following:

Walled-garden data: Perhaps the most important challenge of social media analysis is the limited access to social data. Currently, social media companies keep the data collected from users at their private storage, characterised as private data silos (Figure 95). Third party access to this data is provided for non-commercial purposes through APIs, which enforce rate limiting and/or subsampling of data, and for commercial purposes by directly selling it at a premium price. This creates a competitive advantage for the social media company, which is hard to relinquish, and allows it to enforce its own rules on privacy. It is a challenge for research and society as it stifles independent research and innovation, limits applications of societal value

494 Second Life: https://secondlife.com/
495 K.A. Smith, What's Going on with the Facebook Antitrust Lawsuit? (2021): https://www.forbes.com/advisor/investing/update-facebook-antitrust-lawsuit/
496 B. Auxier, and M. Anderson. "Social media use in 2021." *Pew Research Center* (2021).
497 "Live radio hour in Second Life with Draxtor Despres and Jo Yardley" by HyacintheLuynes: https://commons.wikimedia.org/wiki/File:Second_Life_11th_Birthday_Live_Drax_Files_Radio_Hour.jpg

which do not fit with the companies' business model, and hinders reaching a commonly accepted standard of user privacy. Decentralised social media and increased control of users over their data could be key to overcoming this challenge.

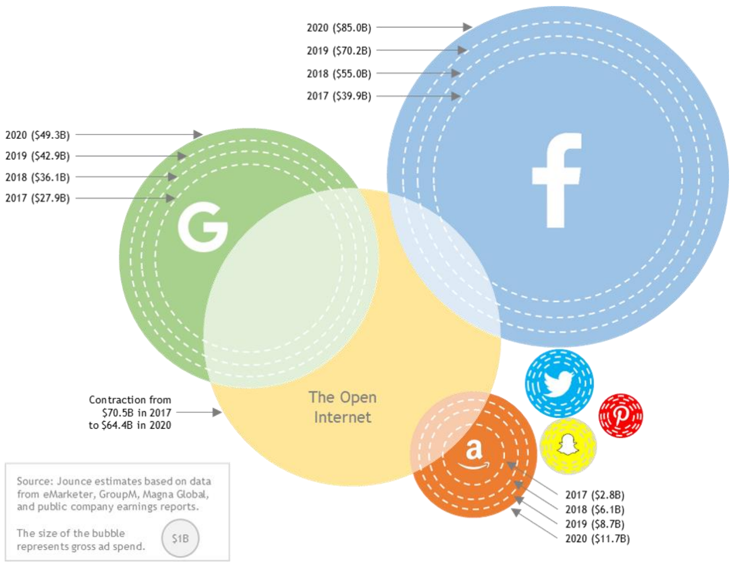


Figure 95: Walled-gardens. The image depicts the total money spent on ads in closed advertising platforms, e.g., by Google, Facebook, and Amazon, compared with open advertising networks ("The Open Internet"). The closed platforms harness first-party data from their users to offer targeted advertising services to their customers[498].

Idiosyncratic content: Social media data, especially text, is highly unstructured and noisy. This is due to the short and casual nature of posts, frequently containing spelling errors, slang, hashtags, emoticons, acronyms and abbreviations in place of detailed descriptions. Crucially, social media posts tend to depend heavily on implied context, potentially related to a niche Internet culture, and frequently contain irony and sarcasm, which is still challenging to detect with NLP algorithms. The above distinguish social from formal text, such as news and encyclopaedia articles, and necessitate extensive cleaning and advanced NLP techniques. These limitations also apply to images and videos, which are frequently processed, e.g., with digital filters that distort the original content.

Need for credibility assessment: Social media contain a lot of rumours, misinformation, and generally content of questionable nature and provenance. A further complication is that social media is often populated by bots that produce both legitimate and illegitimate content, e.g., customer service versus bulk advertising and political propaganda. Automatic credibility assessment at both the content and source level, e.g., through reputation scoring and cross-checking, will be required to use social media data confidently. Spatiotemporal analysis of information propagation on social media graphs is also important to detect false information early and even help mitigate it.

[498] Image from Jounce Media 2020 Market Outlook Report (2020): https://jouncemedia.com/blog/2020/1/22/the-state-of-the-open-internet

Multilinguality and multimodality: Most works on social data analysis focus on a few data modalities, with a strong bias towards English text. The reason is that text is convenient in terms of storage and availability, while English is the dominant language globally. The performance however of AI algorithms can be enhanced with the integration of other data modalities, such as audio, image, and video, as well as multilingual support, in order to understand the large non-English user-base of social media. Currently, multimodal analysis is at an early stage and suffers from a lack of standard datasets. Multilingual support also faces challenges due to the linguistic diversity, breadth of dialects, and lack of training data for minority languages. From a technical perspective, non-English languages cannot be understood by models trained on English and cannot be handled accurately by naive automatic translation. The latter is due to the particularities of syntax, expression, and slang that do not map completely from one language to another. Research on multilingual models, i.e., models that understand many languages simultaneously, has only started emerging. Unfortunately, minority languages have also been overlooked by big social media companies due to the limited opportunities for profit[499,500]. For example, the leaked *Facebook Files*[501], made public in September 2021, have revealed that the company had difficulties in moderating harmful content in non-English developing countries, thus allowing misinformation and hate speech to spread. More effort is thus needed for a linguistically inclusive AI.

Lack of labelled datasets: The success of the DL paradigm was largely fuelled by supervised learning, i.e., algorithms that learn from large-scale annotated datasets. While ML algorithms have achieved marvellous feats, the quality of their output is as good as the data on which they were trained. In the case of social media analysis, data is ample but labelled data is scarce and hard to produce. For example, to create an algorithm that detects depression from user posts, users with actual depression or at least depression symptoms must be identified for ground truth. Acquiring this data requires coordination with domain experts and significant labour, which is time-consuming and expensive. In addition, labels for social issues and social media culture evolve rapidly so that supervised algorithms require frequent updates. This is a key problem of ML algorithms that could be alleviated with stronger transfer learning, unsupervised or semi-supervised algorithms, or even reinforcement learning algorithms that can adapt their learning.

Biased content: Insights from social media are inherently biased for multiple reasons. Firstly, social media users and, specifically, those who post regularly tend to belong to specific demographics, which depend strongly on the examined network. This limits generalisations to the general public. Secondly, social media are prone to amplifying harmful content and fringe views because the employed recommendation algorithms are designed to maximise user engagement and attention, achieved by promoting sensational content. This has led to the creation of echo chambers, i.e., closed communities that reinforce fringe views that end up over-representing them. Finally, social media often suffer from harmful bias and prejudice against

499 N. Nguyen and C. Scurato, Facebook and YouTube's refusal to moderate misinformation in global languages harms communities of color (2021): https://prismreports.org/2021/11/02/facebook-and-youtubes-refusal-to-moderate-misinformation-in-global-languages-harms-communities-of-color/
500 D. Marinescu, Facebook's Content Moderation Language Barrier (2021): https://www.newamerica.org/the-thread/facebooks-content-moderation-language-barrier/
501 Wall street Journal, The Facebook Files: https://www.wsj.com/articles/the-facebook-files-11631713039

minority groups[502,503]. These factors must be considered by AI algorithms to ensure diversity and inclusion instead of bias in their output.

Societal and media industry drivers

Vignette: Multilingual analysis of social media trends, public perception, and sentiment

Danae is a journalist working in a small local news outlet. Danae's daily work includes hunting down important trends and news from social media which may provide the basis and inspiration for Danae's articles. Traditionally, this was done manually, but Danae's company has recently acquired an AI tool that automates this process. Specifically, the AI tool aggregates data from multiple social media and detects trending topics. This has become possible since the recent EU regulation has democratised access to social media data while enforcing common standards of privacy, so that social media analysis can be conducted without exorbitant fees for third-party software. Once the AI tool detects a trending topic, it summarises and categorises it under thematic categories, such as politics, lifestyle, and sports, and presents it to Danae in near real-time or on demand. Danae can choose a topic that piques her curiosity, to see more details, such as the first post from which the topic originated and its propagation in different platforms, presented with an effective visualisation. She can also filter across different locations and time, as well as change the sensitivity of the algorithm so that it can detect more topics more quickly or save cognitive load. The algorithm can detect a variety of topics: yesterday, there were rumours on a fire at Danae's city, while today, there is gossip about the wedding of a local celebrity. Danae loves this tool since it makes her work much easier.

Recently, there have been diplomatic tensions between Danae's country and a foreign one. The relation between the two countries has been historically tricky and the current situation was sparked by a reckless exchange of the two countries' diplomats, which was unusually heated for the diplomatic level. While most news outlets focused on the content of the exchange, fuelling the situation, Danae had the idea to focus on the public perception of the event. With the help of the AI tool, she was able to estimate the public sentiment in the foreign country and found out that the public was overwhelmingly critical of the actions of their diplomat. Key to this discovery was that the AI tool could automatically translate content from foreign social media which Danae would not have understood or would have required manual translation. In addition, the tool helped Danae locate influential and extroverted social media users which were open to interviewing. It then allowed Danae to conduct an immersive interview based on VR technology, which also performed live translation. Finally, Danae gathered data from her own country and wrote a highly influential article that received widespread recognition. Danae's article played a small role in extinguishing the diplomatic tensions between the two countries and had a definite impact on Danae's career.

---

[502] E. Dwoskin, N. Tiku and C. Timberg, Facebook's race-blind practices around hate speech came at the expense of Black users, new documents show (2021): https://www.washingtonpost.com/technology/2021/11/21/facebook-algorithm-biased-race/
[503] V. Vara, Why Do We Care If Facebook Is Biased? (2016): https://www.newyorker.com/business/currency/why-do-we-care-if-facebook-is-biased

Future trends for the media sector

As the web becomes more and more social, the importance of social network analysis and automatic extraction of intelligence from social data will grow. The media sector will need to ***extract complex information from social media that go beyond the current practices***, for example, estimating health indexes of the environment, cultural heritage, and mental health and tracking down misinformation networks automatically. In the latter case, journalism will increasingly rely on tools that automatically detect bots and distinguish fake from legitimate content in social media, by assessing the veracity and source provenance of information. The developed technology will be responsible for gathering information from diverse sources and presenting them in a balanced unbiased way to the journalists, thus helping them bypass their personal biases. Ultimately, this will increase the quality of information offered to the end users and citizens (more information on the topic of disinformation can be found in section 9.4)

In brief, key future trends include:

Decentralisation and democratisation of social content: Current efforts to decentralise the Web are expected to carry over to social media, empowering users with control over their personal data while simultaneously opening them to a larger audience. Online social interactions will also generalise across online applications and may eventually become public. This data will present great opportunities for analysis and will be valuable to the media sector.

Privacy and security: User privacy and security are existing trends that will grow in importance with modern applications that rely heavily on user data. Processing this data while simultaneously minimising exposure to unwarranted parties is a key challenge that future applications will need to address. An additional challenge, considering the trend to adopt unique and authentic identities online is identity theft and impersonation. Candidate solutions include advanced encryption, anonymisation, obfuscation, and differential privacy techniques all enabled by AI.

Multimodal, multilingual, and multidomain analysis: Current AI applications on social data rely heavily on English text from a single social media, typically Twitter, but this is rapidly changing. Aggregating information from multiple languages, modalities, and social domains is a natural evolution with clear benefits to the accuracy and diversity of social media analysis.

Credibility assessment algorithms: As misinformation and low content quality is typical of social media, the importance of automatic filtering, cross-referencing, source provenance detection and authoritativeness estimation will grow. These tasks will rely heavily on advanced processing on dynamic graphs and information fusion from multiple sources.

Addressing bias: As the harmful impact of bias is widely recognised and humans become increasingly aware of their susceptibility to it, there is a pressing need for automatic detection and mitigation. This need has sparked fruitful research and resources on the operationalisation of bias, which is expected to grow. The trend is important for the media industry, as the latter needs to abide by strict measures of objectivity, balanced news reporting, and journalistic ethics.

Increased immersion: The recent success of VR and AR technologies has affected social media and the announcement of the Metaverse indicates that it is an important trend. Online social interactions are thus expected to become more immersive, leading to new types of social content and applications, which can be mined by ML algorithms for novel insights. At the same time, immersion carries increased exposure and new privacy concerns that must be addressed to ensure the widespread adoption of the technology. Key opportunities for AI include the generation of avatars, landscapes, and content to populate this world, the development of sentient AI agents and affective technologies to make the interactions between humans and AI agents more human-like, and the removal of language barriers between metaverse users.

Goals for next 10 or 20 years

Based on the aforementioned trends, we can extrapolate in the near future and predict the following goals:

Open, inclusive, and safe social web: Tomorrow's web will be distinctly more social. People will be authenticated by their unique online ID and will operate similarly as they would in the physical world. Proper authentication will result in safer and socially responsible conducts, while citizens will be protected from digital identity theft. Social interactions will abound and public data will be readily available for analysis, subject to user defined privacy constraints. In this world, both media and citizens will have access and benefit from insights resulting from the analysis of open social data with advanced AI techniques.

Accurate social media analysis: Based on the availability of social media data, advanced processing, and information fusion from multiple sources, social analytics will be even more accurate and useful to stakeholders, including the media industry. Journalists will have timely access to social data and ask complex questions that are impossible at present. The answers will be insulated from misinformation and bias thanks to the advanced AI algorithms but also thanks to human education and positive change in how society thinks.

Metaverse: Following the trend of increasing immersion, next generation social media will attain the vision of metaverse. Online profiles will be substituted by 3D-rendered avatars that will function as their digital twins in the virtual world. People will be able to customise their avatars as well as their surroundings with imaginative content created with the help of AI technologies. People will also participate in immersive experiences such as convening in digital cafes, attending digital concerts, and playing holographic games, and have meaningful interactions with other users as well as AI agents, which will successfully mimic human behaviour. Communication barriers will be removed with the help of advanced language processing. The virtual world will bring a serious paradigm shift and will change the way we think about the world and the media.

9.6 AI for entertainment and film production

Contributors*: Lorenzo Seidenari (UNIFI), Federico Becattini (UNIFI), Marco Bertini (UNIFI)*

Current status

The production of feature movies and other entertainment content like TV series is extremely complex and composed of very diverse steps that require different types of optimisations to reduce their costs and increase the chance of success. During the last years, several techniques based on AI have been proposed for the film and TV industry, addressing the needs of pre-production (e.g. to reduce the cost of animatics storyboarding), production (e.g. via virtual cinematography), post-production (e.g. via the first deepfakes), screening and distribution (e.g. via recommendation algorithms and audience analysis). Below, we report a few notable applications that are emerging and are currently starting to be exploited in cinematography.

Currently, companies offering a complete solution[504] to support film production are also able to provide automatic storyboarding. Given a script and a dataset of already shot footage, these systems are trained to provide suggestions to directors, comprehensive of shot size, type and description (Figure 96).

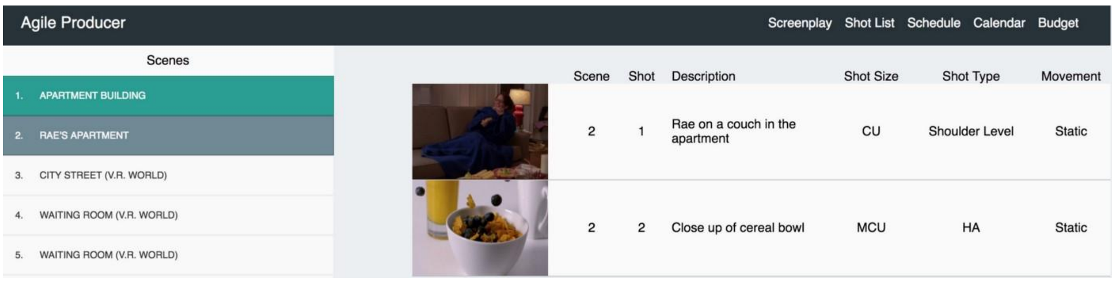


Figure 96: Automatic Storyboarding tool based on AI (RivetAI). [504,505]

Moreover, current large scale language models[506] can create synthetic text either from scratch or by completion[507]. This kind of capability can provide a strong support in creation of dialogs and scripts. This kind of large-scale architectures are also able to perform machine translation tasks, given paired sets of sentences[508]. Automatic translation via Deep Learning will empower the movie industry, allowing to target the global market easily.

During the last decade thanks to diffusion and low cost equipment such as drones and 360° cameras, AI methods have been deployed to allow a certain degree of automation in the capture of video content. At the same time, deep learning techniques have matured to the point of being able to generate realistic synthetic content in different media, e.g., video and audio.

Immersive media is also gaining popularity, particularly after the revelation of the so-called Metaverse. However, bandwidth and latency limits hinder the online experience of immersive

---

504 RivetAI: https://www.rivetai.com/

505 Image source: D. Ray, Data Science and AI in Film Production (2018), https://medium.com/rivetai/data-science-and-ai-in-film-production-8918ea654670

506 OpenAI, GPT-3 Powers the Next Generation of Apps (2021): https://openai.com/blog/gpt-3-apps/

507 Jasper - The Future Of Writing: https://www.jasper.ai/

508 S. Edunov, M. Ott, M. Auli, and D. Grangier. 2018. Understanding Back-Translation at Scale. In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing, pp. 489–500, Brussels, Belgium. Association for Computational Linguistics.

media such as Virtual Reality (VR). 360° videos, for example, that are meant to be viewed on a Head-Mounted Display (HMD), require data rates that are around two orders of magnitude higher than standard videos to provide the same quality impression. Distributing improved quality levels in the user's Field of View (FoV) is a simple way to reduce the required data rates. When streaming, this demands anticipating where the user will look, which can be done as far ahead of time as a few seconds if needed.

AI is addressing movie production challenges in all modalities. Current state-of-the art voice generators can provide industry grade voiceovers with hyper realistic synthetic voices[509]. This allows for a great reduction in film budget by not requiring to hire multiple actors to create a multilingual product. Moreover, if this technology is combined with automated translation, the market reach of movie producers is practically global.

Recently, Neural Radiance Fields (NeRF)[510] have revolutionised the field of novel view synthesis from single images. Video generation methods allow to create new video content from some original content, changing camera views, content style and animating differently the persons framed in a scene, as shown in Figure 97.

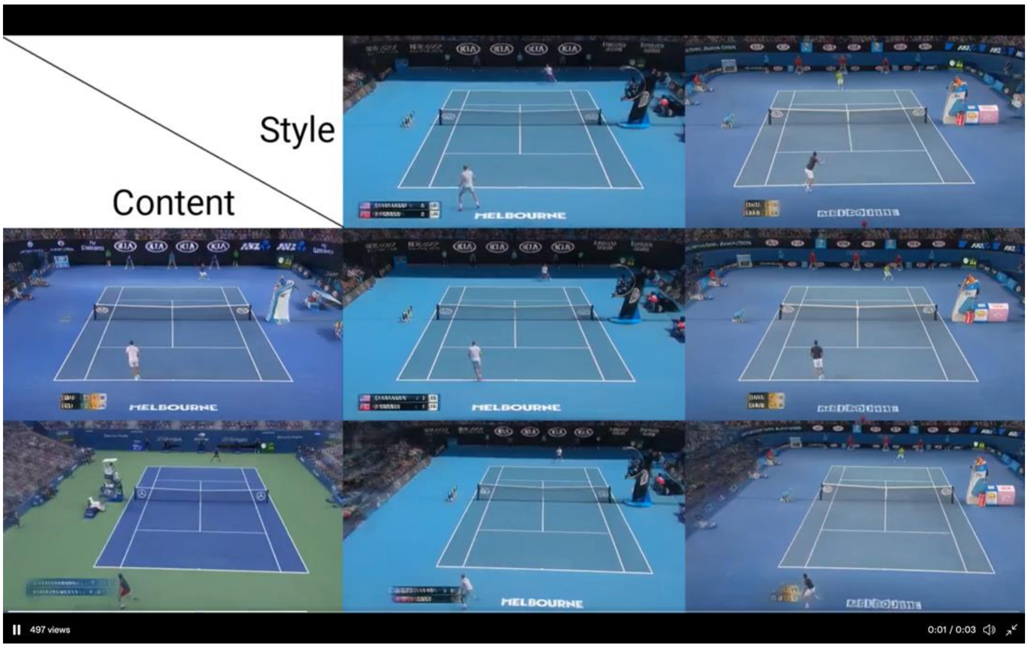


Figure 97: Example of synthetic video generation[511].

[509] Murf AI voice generator: https://murf.ai/

[510] B. Mildenhall, P. P. Srinivasan, M. Tancik, J. T. Barron, R. Ramamoorthi, Ren Ng, "Nerf: Representing scenes as neural radiance fields for view synthesis", Communications of the ACM, January 2022, Vol. 65 No. 1, pp. 99-106

[511] Image source: W. Menapace, A. Siarohin, C. Theobalt, V. Golyanik, S. Tulyakov, S. Lathuilière, E. Ricci, "Playable Environments: Video Manipulation in Space and Time", 2022, https://arxiv.org/abs/2203.01914

Also recently, thanks to AI enabled robotic pipelines automatic shooting of complex videos has become a reality (Figure 98).[512] Multiple techniques are combined allowing one or more drones to execute valid camera movements and film scenes in novel ways. Autonomous aerial cinematography has the potential to enable automatic capture of aesthetically pleasing videos without requiring human intervention.

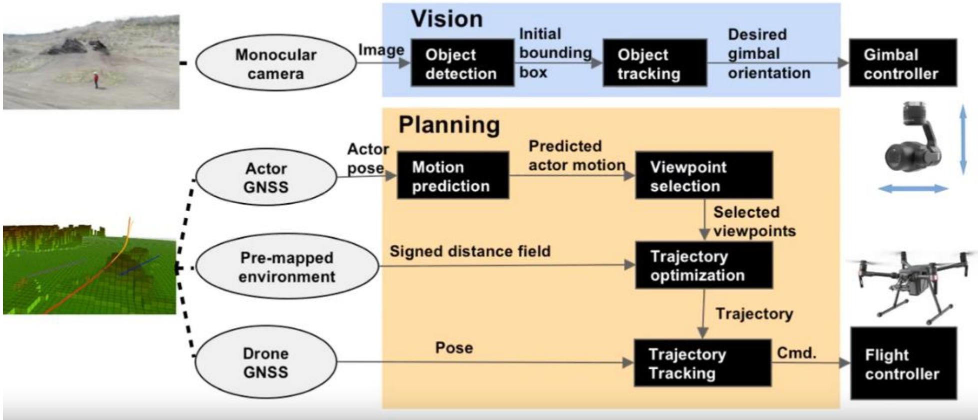


Figure 98: Example of an AI drone video filming system[513].

Deepfake technology seems to be one of the most disruptive technologies in Hollywood and it will probably drastically change the whole worldwide movie industry in the next few years. One of the first possible applications could be in film production based on the use of de-aging technology. In fact, it will be possible both to bring back performers who are no longer alive or even achieve realistic depictions of their younger selves. This opens new and exciting opportunities for film making, advertisements, historical documentaries and so on. As seen in Figure 99, deepfake technology is mature and allows face replacement, thus offering a great opportunity to, for example, reshoot scenes or work with stunt doubles in dangerous or acrobatic film sequences.

Finally, regarding content distribution, acquisition and marketing, more and more companies rely on machine learning to perform personalisation and recommendation of content, helping users find content that satisfies their needs or current mood. Thanks to the large number of customers that streaming industry leaders like Netflix and Amazon have, such recommender systems are always fed with a large set of high-quality annotations. As an example, it is estimated that 80% of the shows watched in Netflix are found through the company's recommender system[514]. This system is fed with various data, including, interactions with the website (browsing history and ratings), relationships and similarities between the preferences

[512] R. Bonatti, Y. Zhang, S. Choudhury, W. Wang, S. Scherer, Autonomous drone cinematographer: Using artistic principles to create smooth, safe, occlusion-free trajectories for aerial filming (2018), https://arxiv.org/abs/1808.09563

[513] Image source: YouTube (R. Bonatti, Autonomous drone cinematographer, ISER 2018) - https://www.youtube.com/watch?v=QX73nBBwd28

[514] A. Krysik, Netflix Algorithm: Everything You Need to Know About the Recommendation System of the Most Popular Streaming Portal (2021): https://recostream.com/blog/recommendation-system-netflix

of individual user groups, and information about the content itself (genre, language, actors, etc.).[514] To offer a more personalised experience, additional training data include the time of the day when individual users use the service, type of device used to view content, and average viewing length[514].

Figure 99: Example of deepfakes posted on social media[515].

Research challenges

One of the main challenges we can identify across most of the aforementioned approaches is linked with **data scarcity**. This is limiting approaches such as machine translation or speech generation to a few languages for which such aligned datasets exist. Unfortunately, this hinders the actual advantage brought by these approaches, which is the capability to post-produce movies in a large number of languages.

Video generation methods, on the other hand, are currently limited in their capabilities of content creation given that they can animate persons only in **relatively static environments**, without handling interactions with objects in the background.

Automated cinematography methodologies are still in their infancy and are currently reliable only when dealing with a **few manoeuvres**. Moreover, such methods are only applicable to camera motions derived from UAVs flying at an altitude. A lot of shooting happens in a much diversified manner, using for example dolly cameras, or handheld cameras. More complex systems and agents should be developed in order to cope with these situations.

Compressing content either in VR setups or standard streams in a way that the viewer finds pleasant without wasting bandwidth is still an unsolved challenge. Another issue that is not yet solved is the capability of codecs to be personalised for a single viewer. While this is somehow the main idea behind the delivery of 360° videos, it is extremely challenging for standard video streams. All in all, we are not yet able to characterise correctly viewer behaviour individually.

[515] Image source: TikTok (@deeptomcruise): https://www.tiktok.com/@deeptomcruise?lang=en

High resolution content generation from textual descriptions still produces images with a quality and resolution that do not allow to use them in production. Generation of backgrounds and settings for virtual set filming is currently still mostly performed using traditional CGI methods. Developing generic image generators that have the same quality of those developed for limited domains like faces is still a challenge for the scientific community.

Regarding personalisation of content, it is still to be debated if personal preference can be derived directly from multimodal cues. We are still not able to build a model of personal likings directly from content.

Societal and media industry drivers

Vignette: Facilitating film shooting using drones and enlarging reach of content via automatic translation and dubbing

Jane is a video producer that works with a low budget. She employs innovative methodologies for filming such as drones and 360° cameras so that she can work with a limited filming crew. At the same time, she would like to extend her business and distribute content on more countries.

Recently, Jane has started to invest on AI based products that allow to speed up her filming process. Instead of requiring multiple drone pilots, Jane acquired a software to automatically shot video sequences from drones with minimal supervision. The software allows to instruct drones to follow specific subjects or to perform smart manoeuvres in order to shoot cinematographically sensible sequences.

To improve the reach of her content, she also acquired an AI based automatic dubbing and translation tool. Thanks to speech synthesis, subjects can be dubbed at low cost. Voice pitch is differentiated thanks to computer vision that automatically associates the same synthetic voice actor to the same person. Jane can still control and decide the voice pitch in order to obtain a product that suits her needs. Together with automatic voice generation, the AI tool also includes a speech recognition system and an automatic translator, allowing to extract speech from original sequences, create subtitles automatically as well as translations in multiple languages.

During the editing phase of her latest video, Jane notices that unluckily one of the scenes has not been filmed from the best point of view. Filming it again, however, is too expensive for her. Using a new tool for scene generation, Jane picks some frames from the filmed scene and recreates an animated virtual view that she can use as if it was filmed during the production.

Future trends for the media sector

In the following, we briefly summarise some the ways in which AI is expected to enhance and facilitate film production, from the pre-production stage to content distribution:

- Comprehensive **tools to organise and direct drone swarms**, aiming to improve the shooting process by filming aesthetically pleasing landscapes, action-packed film scenes or military action in war zones from novel points of view and in settings that would be difficult or dangerous for human camera operators to be.

- Comprehensive **tools to perform voice recognition, synchronisation, translation and dubbing**, aiming to address the needs of different international markets and provide localised contents.
- **Automatic content creation** (e.g. trailers for different audiences or personalised trailers, movie/TV scenarios, CGI, music to match scenes/emotion/dialogue, etc.). These techniques will significantly facilitate production by reducing costs of virtual sets, post-production through virtual cinematography and automated VFX, and marketing and promotion through targeting of demographics of clients using automated and personalised trailers.
- **Deep fakes for movies**. The current initial use of this technology aims to replace actors, e.g. to depict a younger appearance of an aged actor in a film sequel. Extension to full scenes and settings can be expected in the next few years.
- AI for **audience analysis** to provide insights on what scenarios have the potential to be popular movies, which actors to cast, etc. Automated and multimodal AI test screening allows to find the best market and demographics for each title and also to make decisions about casting or the selection of scripts, by analysing the emotions and reactions of audiences without requiring to organise a costly in-person event.
- **Automation of film directing/editing/shooting processes**. Through automatic selection of the best filmed scenes, virtual cinematography allows to create new scenes from the filmed ones while automated camera movement and tracking improve the shooting process.
- **Film aesthetics and style transfer**. Creating different types of visualisations from a common source, e.g. transforming movies into cartoons, automatic colour grading aimed at inducing different emotions, etc., may open new avenues in film production.
- **Interactive and personalised movies**. Using AI to extend certain parts and subplots of a movie with materials that are more interesting for certain audiences, or creating personalised experiences based on the reactions of the audiences (as measured e.g. by wearable sensors or cameras) can lead to innovative ways of experiencing cinema and TV in the near future.

Goals for next 10 or 20 years

AI applications are expected to become pervasive in cinematography and media content production in the next few years. A goal is to exploit the diffusion of Unmanned Aerial Vehicles (UAVs, or drones) to automatise shooting, aiming not only to facilitate the filming process but also to optimise viewer experience. In fact, we expect drone manoeuvres to be programmable based on the movie style or the emotion that the filmmaker wants to communicate. We also expect the AI piloting the drone to automatically infer such emotion from the observed scene and adapt its trajectory to convey an adequate shooting style, e.g., fast close-ups for dynamic action-packed scenes or slow movements from a distance for sentimental scenes.

Another goal is to make UAVs smarter. Onboard AI could entail a high-level understanding of the shot that could be used to enhance the quality of the footage by optimising framing position or dynamically tracking different targets depending on their estimated importance.

Another interesting application would be to shoot different versions of the same content with multiple drones that adhere to a set of shooting styles. Such diversity could be used in a post-processing step by a human or AI operator to build a rich footage or could be directly provided to users, offering different experiences depending on their personal tastes and preferences. Furthermore, drones could also be equipped with the capability of predicting trajectories of relevant subjects or objects. By being able to forecast motion patterns in the observed scene, we expect drones to anticipate movements and prepare their shooting position in advance to obtain more spectacular and memorable scenes.

Similarly to trajectory prediction, forecasting human gaze could also play an interesting role both in content generation and fruition. In fact, drones are often piloted by human operators using 360° headsets. Predicting their gaze could allow the drone to anticipate the actual command given by the operator and execute smoother manoeuvres. We expect the same kind of technology to be useful for watching 360° video content, especially when streamed online. Predicting user gaze can allow to adaptively compress the video stream, considerably reducing bandwidth. 360° content will become increasingly pervasive, also affecting how content will be produced. Videos customised for headsets in fact will open a whole new range of opportunities for content creators. For instance, videos could be shot to contain multiple narrations depending on which portion of the video is observed by the user throughout the viewing experience.

Now, it is not possible to imagine all the infinite opportunities where deepfake technology could be used besides obtaining amazing and unfeasible results such as shooting a new movie with a not-living actor. Deepfakes could be adopted to strongly reduce production costs. In fact, such a technology would also be able to replace and/or adapt the original performance for a different goal without the need of getting an entire set and crew assembled to change a part of a dialogue or to repeat a movie scene again. For instance, some dialogues and scenes could be recorded with a reduced part of the cast and with some proper stunt doubles (see Figure 99 above) and be successively completed by resorting to deepfake technologies to modify faces, facial expressions, lips and eye movements accordingly. It is easy to understand how much deepfake technologies which basically deal with faces could become crucial in post-processing operations in order to change and modify what has been recorded without reshooting. For example, it could be used by filmmakers to manipulate and alter specific dialogues without needing to do reshoots.

9.7 AI for games

***Contributors**: David Melhart (modl.ai), Ahmed Khalifa (modl.ai), Daniele Gravina (modl.ai)*

Current status

During the last decade ***game AI*** as a research area has been becoming more and more popular in both academic and industrial circles. This is evidenced by increasing interest in AI by industry stakeholders and the large volume of academic studies, which has been supported by an active and healthy research community — roots of which point well beyond the past decade to at least since the start of the IEEE CIG and the AIIDE conference series in 2005. While initially most of the work published at academic venues was concerned with learning to play a particular game or using search/planning algorithms to play a game without learning, gradually, a number of new applications for AI in games and for games have come to complement the original focus on AI for playing games[516,517].

Since the early days of the field, papers on procedural content generation, player modelling, game data mining, human-like playing behaviour, automatic game testing and so on have become commonplace within the community. While academic interest in game AI applications has been on the rise in the last decade, industry and academia do not necessarily attempt to solve the same problems with the same approaches. Nevertheless, it may be that more traditional algorithmic solutions emerging from industry can inspire new approaches in academia and vice versa. Consequently, what we see today is a healthy indication of a parallel progress with a certain degree of collaboration.

Figure 100: The increase in the game development cost over the past four decades[518].

While it can be argued that ***Non-Player Characters*** (NPC) have been solved to a satisfactory degree with sophisticated behaviour trees, game AI research can offer much more to the industry. The multidisciplinary nature of game AI and a more pragmatic and holistic view of the game AI problem have shifted academic and industrial interests in recent years. It seems that we have long reached an era where the primary focus of the application of AI in the domain of games is not on agents and NPC behaviours. The focus has, instead, started to shift towards interweaving game design and game technology by viewing the role of AI holistically and integrating aspects of ***procedural content generation*** and ***player modelling*** within the very

516 G. N. Yannakakis. Game AI revisited. In Proceedings of the 9th conference on Computing Frontiers, pages 285–292. ACM, 2012

517 G. N. Yannakakis, and J. Togelius. *Artificial intelligence and games*. Springer, 2018.

518 Image source: R. Koster, VentureBeat, The cost of games (2018) https://venturebeat.com/2018/01/23/the-cost-of-games/

notion of game AI[78]. While AI can help to make better games, this does not necessarily happen through better, more human-like or believable NPCs. Notable examples of non-NPC AI in games include *No Man's Sky* (Hello Games, 2016) and its procedural generation of a quintillion different planets, *Nevermind* (Flying Mollusk, 2016) with its affective-based game adaptation via a multitude of physiological sensors, and – more contemporarily – *Watch Dogs Legion* (Ubisoft, 2020) with procedurally generated characters and missions. But there might be other AI roles with game design and game development that are still to be found by AI.

Beyond playing games and content generation, AI might be able to play the role of a ***design assistant***, a ***data analyst***, a ***playtester***, a ***game critic***, or even a ***game director***. Large industry players like Ubisoft and King are working on integrating AI into their data analytics[519], content recommendation[520] and moderation[521], player modelling[522], and level design[523] pipelines in games like *Tom Clancy's The Division 1-2* (Ubisoft, 2016; 2019), *For Honor* (Ubisoft, 2017; 2018), *Candy Crush Saga* (King, 2012) and *Candy Crush Soda Saga* (King, 2014). Indeed the AI assistance in data processing, content generation, testing, and moderation is becoming more and more sought-after by industry stakeholders as new "game as a service" business models turn game development and maintenance into an ongoing and costly process. Companies like modl.ai are aiming to meet the demand with sophisticated AI tools that aid crucial parts of these processes.

Finally, beyond the games industry, video games are still used as benchmarks for more complex AI architectures. In recent years, we saw tech giants like OpenAI and the Google-affiliated DeepMind test their algorithms on esports games. As great milestones in both game AI and AI in general, OpenAI's *OpenAI Five* became the first AI system to defeat the world champions in *Dota 2* (Valve Corporation, 2013)[524] and DeepMind's *AlphaStar* was able to perform at a grand-master level in *StarCraft II* (Blizzard Entertainment, 2010)[525]. While advancements can be considered great stepping stones towards more complex general AI, they show the strength of games as benchmarks for AI applications beyond the games industry.

Research challenges

There are a number of challenges still open in different fields of game research. The intersection of key areas such as game playing, content generation, and player modelling hide numerous pitfalls when we are considering combining these methods to achieve more complex goals (e.g. combining game playing and player modelling in AI-assisted automated testing). While NPC AI

[519] A, Canossa, et al. "Like a DNA string: Sequence-based player profiling in Tom Clancy's the Division." Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment. Vol. 14. No. 1. 2018.
[520] L. Cao, et al. "Debiasing Few-Shot Recommendation in Mobile Games." ORSUM@ RecSys. 2020.
[521] A. Canossa, et al. "For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay." Proceedings of the ACM on Human-Computer Interaction (CHI PLAY), 2021.
[522] D. Melhart, et al. "Your gameplay says it all: modelling motivation in Tom Clancy's The Division." 2019 IEEE Conference on Games (CoG). IEEE, 2019.
[523] V. Volz, et al. "Capturing local and global patterns in procedural content generation via machine learning." 2020 IEEE Conference on Games (CoG). IEEE, 2020.
[524] C. Berner, et al. "Dota 2 with large scale deep reinforcement learning." arXiv preprint arXiv:1912.06680, 2019.
[525] O. Vinyals, et al. "Grandmaster level in StarCraft II using multi-agent reinforcement learning." Nature 575.7782, 2019.

might be a solved issue for gameplay purposes as far as the industry is concerned, game-playing agents that test games or replace human players still hold value to stakeholders. While research often focuses on limited testbeds and single-use applications, a great challenge lies in scaling these pipelines to production environments.

Game playing is one of the core research areas of game AI. Even though today's landscape is more diverse in terms of research topics, game playing still takes up the majority of published papers in the field[526]. While early studies and even recent high-profile examples (see DeepMind and OpenAI above) are focusing on playing to win (Figure 101), introducing other goals into the system is not straightforward. The main areas which can intersect game playing are playing for experience, player modelling, and Procedural Content Generation (PCG).

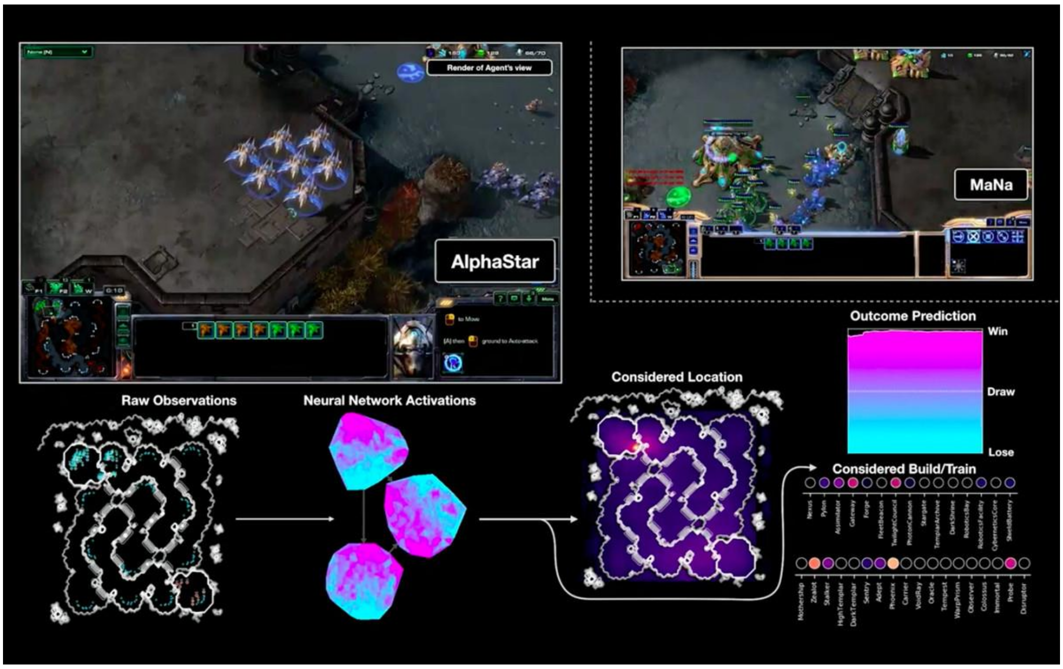


Figure 101: Visual representation of the AlphaStar agent while it plays against human champion MaNa. The agent sees the game map and predicts what actions it should make to lead it to victory [527].

In the intersection of playing to win, playing for experience, and player modelling the focus is on ***believable play***. There is a direct link between player modelling and believable agents, as research carried out for the modelling of human, human-like, and supposedly believable playing behaviour can inform the construction of more appropriate models for players. However, an agent cannot be believable or existent to augment the game's experience, if it is not proficient. Being able to play a game well is in several ways a precondition for playing games in a believable manner. Nevertheless, player models can inform and update believable agent architectures. ***Models of behavioural, affective and cognitive aspects of gameplay*** can improve the human-

[526] G. N. Yannakakis, and J. Togelius. *Artificial intelligence and games*. Springer, 2018.
[527] Image source: DeepMind - https://deepmind.com/blog/article/alphastar-mastering-real-time-strategy-game-starcraft-ii

likeness and believability of any agent controller—whether it is ad-hoc designed or built on data derived from gameplay. While the link between player modelling and believable agent design is obvious and direct, research efforts towards this integration within games are still sparse. However, the few efforts made on the imitation of human game playing for the construction of believable architectures have resulted in successful outcomes.

Procedural Content Generation (PCG) is one of the areas of recent academic research on AI in games, which bears the most promise for incorporation into commercial games. A number of recent games have been based heavily on PCG, including independent ("indie") game production successes such as *Spelunky* (Mossmouth, 2009) and Minecraft (Mojang, 2011), and mainstream AAA games such as *Diablo III* (Blizzard Entertainment, 2012), *Civilization V* (2K Games, 2010), *No Man's Sky* (Hello Games, 2016), Borderlands (Gearbox Software, 2009), and *Watch Dogs Legion* (Ubisoft, 2020) (see Figure 102). Some games heavily based on PCG and developed by researchers have been released as commercial games on platforms such as Steam and Facebook. Where PCG intersects game playing we can see agents that are capable of playing a game proficiently, which can be useful for ***simulation-based testing***, i.e., the testing of newly generated game content by playing through that content with an agent (see section 5.1). Moreover, if an agent is trained to perform well in only a single game environment, it is easy to overspecialise the training and arrive at a policy/behaviour that will not generalise to other levels. Therefore, it is important to have a large number of environments available for training. PCG can help with this, potentially providing an infinite supply of test environments for different agents.

Figure 102: Procedural Content Generation is used for the generation of maps in Civilization games[528].

[528] Image source: The Scientific Gamer - https://scientificgamer.com/the-procedural-generation/

Societal and media industry drivers

Vignette: AI for content moderation in multiplayer games

Ella is a content moderator on a new multiplayer team shooter game. Ella's job is to investigate and evaluate cases of in-game harassment and cheating. In the past, these instances were investigated based on tickets submitted by players (see Figure 103), which made the process slow and unreliable. On one hand, an open ticket system was prone to abuse, while on the other, many victims didn't report harassment or cheating but stopped playing the game altogether instead. However, today Ella is working with automated AI tools, which analyze in-game behaviour, game telemetry, and chat logs to flag problematic interactions. The new system is able to identify not just severe cases of harassment but more subtle toxic behaviour, which used to go mostly unreported in the past. While the AI tool is not banning people from the game automatically, it alleviates much of the strain of having to look through gameplay logs manually, allowing Ella to focus on more crucial parts of the investigation. Additionally to identifying toxic behaviour, the tool Ella uses also flags potential instances of cheating and abuse of game mechanics. To retain players – beyond catching toxic behaviour – it is especially important in multiplayer games to minimise cheating and exploits. While cheaters can be reprimanded and banned if necessary, it is also important to identify exploits that can be patched out in later content updates. This way the AI-assisted monitoring of the game aids not just Ella's work as a content moderator but her designer and developer colleagues' as well.

Figure 103: Traditional in-game reporting tool for players in For Honor (Ubisoft, 2017)[529].

Future trends for the media sector

As highlighted in the previous sections, papers on procedural content generation, player modelling, game data mining, human-like playing behaviour, automatic game testing, and so on have become commonplace within the community, and industry adaptation of these methods are already on their way. However, almost all research projects in the game AI field are very specific. Most published papers describe a particular method—or a comparison of two or more methods—for performing a single task (playing, modelling, generating, etc.) in a single game.

[529] A. Canossa, et al. "For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay." Proceedings of the ACM on Human-Computer Interaction (CHI PLAY), 2021.

This is problematic in several ways, both for the scientific value and for the practical applicability of the methods developed and studies made in the field. If an AI approach is only tested on a single task for a single game, how can we argue that is an advance in the scientific study of artificial intelligence? And how can we argue that it is a useful method for a game designer or developer, who is likely working on a completely different game than the one the method was tested on?

The focus of generality solely on play is very narrow as the possible roles of AI and general intelligence in games are many, including game design, content design, and player experience design. The richness of the cognitive skills and affective processes required to successfully complete these tasks has so far been largely ignored by game AI research. We thus argue, that while the focus on general AI needs to be retained, research on ***general game AI*** needs to expand beyond mere game playing. The new scope for general game AI beyond game-playing broadens the applicability and capacity of AI algorithms and our understanding of intelligence as tested in a creative domain that interweaves problem-solving, art, and engineering.

For general game AI to eventually be truly general, we argue that we need to ***extend the generality of general game playing*** to all other ways in which AI is (or can be) applied to games. We should develop methods that can model, respond to, and/or reproduce the very large variability among humans in design style, playing style, preferences, and abilities. This generality can be embodied in the concept of general game design, which can be thought of as a final frontier of AI research within games[530]. It is important to note that we are not arguing that more focused investigations into methods for single tasks in single games are useless; these are often important as proofs-of-concept or industrial applications and they will continue to be important in the future, but there will be an increasing need to validate such case studies in a more general context. We are also not envisioning that everyone will suddenly start working on general methods. Rather, we are positing generalisations as a long-term goal for our entire research community.

Finally, the general systems of game AI that we envision ought to have a ***real-world use***. There is a risk that by making systems too general we might end up not finding applications of these general systems to any specific real-world problem. Thus, the system's applicability (or usefulness) sets our core constraint towards this vision of general game AI. More specifically, we envision general game AI systems that are nevertheless integrated successfully within specific game platforms or game engines.

Goals for next 10 or 20 years

The shift of focus towards generalised AI solutions is a trend that is already taking place in other fields of AI research. So-called *foundation models* are on the rise in areas such as natural language processing and computer vision. These models are large pre-trained networks leveraging deep learning and transfer learning to provide easily adaptable AI for broad applications. It is not surprising that many of these models are the driving force behind

[530] J. Togelius and G. N. Yannakakis. General General Game AI. In 2016 IEEE Conference on Computational Intelligence and Games (CIG). IEEE, 2016.

innovation in many industries. While ***foundation models for games and gameplay*** are not existing yet, seeing the success of these models in other fields, the path seems clear. In the future, we are expecting to see both the more widespread integration of existing foundation models into games and the rise of ***game-based models*** in areas such as level design, believable game playing, and player modelling. The adaptation of existing models has already begun with games such as *AI Dungeon* (Latitude, 2019), which incorporates GPT-2 and GPT-3 foundational language models.

In the next decades, we expect to see innovations in different areas of the games industry driven by these foundation models. As the shift towards a "games as a service" model continues, the need for testing new content will exceed human capacity. ***Testing games*** for bugs, ***balancing player experience and behaviour***, and other issues is important in game development, and one of the areas where game developers are already looking for AI assistance. In the future, we expect to see the rise of foundation models which are built to navigate ad-hoc game spaces, test puzzles, or provide generalised feedback of player states. For the particular case of finding bugs and exploits in games, one of the research challenges is to find a good and representative coverage of problems, so as to deliver an accurate picture to the development team of how many problems there are and how easy they are to run into and allow prioritisation of which problems to fix.

More than a decade ago, the outstanding feature of *Left 4 Dead* (Valve Corporation, 2008) was its ***AI director***, which adjusted the onslaught of zombies to provide a dramatic challenge curve for players. While simple and literally a single dimension of player experience was tracked, the AI director proved highly effective. In the future, new advancements in deep learning will allow for much room for creating more sophisticated AI directors; the experience-driven PCG framework[531] is one potential way within which to work towards this.

We also expect to see the rise of ***AI-Based Game Design***. This could be seen as an opportunity to showcase AI methods in the context of games, but it could also be seen as a way of advancing game design. Most classic game designs originate in an era where there were few effective AI algorithms, there was little knowledge among game designers about those AI algorithms that existed, and CPU and memory capacity of home computers was too limited to allow anything beyond simple heuristic AI and some best-first search to be used. One could even say that many classic video game designs are an attempt to design around the lack of AI—for example, the lack of good dialog AI for NPCs led to the use of dialog trees, the lack of AIs that could play First-person shooter (FPS) games believably and competently led to FPS game designs where most enemies are only on-screen for a few seconds so that you do not notice their lack of smarts, and the lack of level generation methods that guaranteed balance and playability led to game designs where levels did not need to be completable. The persistence of such design patterns may be responsible for the relatively low utilisation of interesting AI methods within commercial game development. Advancements in computation and AI optimisation in the next decades will certainly make AI methods far more accessible in the future. By starting with AI and designing a

[531] G. N. Yannakakis and J. Togelius. Experience-driven procedural content generation. Affective Computing, IEEE Transactions on, 2(3):147–161, 2011.

game around it, new design patterns that actually exploit some of the recent AI advances can be found.

Finally, as AI becomes ***more ubiquitous and easily adaptable*** in the future, we expect to see the rise of ***AI-Assisted Game Development***. By combining the AI techniques mentioned above, new co-creation algorithms will be able to aid developers in providing faster iteration times and greater customisation in content creation. Beyond leveraging computational creativity to create new content and various AI methods to test on the fly, the future will most probably see a push for more personalised content as foundational player models are becoming more widely used. Future AI applications might be able to create new content or fine-tune the experience of any game for any emotional state, completing a meaningful affective loop, which defines a framework that is able to successfully elicit, detect and respond to the cognitive, behavioural and emotive patterns of the players[532].

9.8 AI for music

Contributors: *Rémi Mignot (IRCAM), Axel Roebel (IRCAM)*

Current status

In the early stages of ***automatic music processing***, such as in musical representation or sound signal modelling, ad hoc methods and expert algorithms were explicitly designed by the researchers or engineers to achieve specific tasks. The targeted applications were for example: tempo estimation, harmonic key recognition, automatic transcription, audio indexing, similarity estimation, automatic mixing, beat and downbeat detection; for which the experts of the field are able to formalise explicitly the models for the properties to analyze. Nevertheless, these approaches were not sufficient to explain more complex concepts of the music, such as the ***musical genre or the emotion*** provided by a song. Therefore, since almost two decades, machine learning has been used to build models able to automatically learn a musical concept, without having to formalise it mathematically. Frequently inspired by automatic speech recognition, these methods have been able to solve more complex tasks based on a dataset of examples, such as the musical genre recognition[533].

For some years now, deep learning has made possible to go even further. Many ad hoc methods, which worked fine, have been overtaken by deep neural network models. For example, whereas previous tempo estimation models were based on a dedicated signal transformation, pre-established rhythmic patterns and pattern recognition methods[534], more recent approaches are

[532] P. Sundstrom. Exploring the affective loop. PhD thesis, Stockholm University, 2005.
[533] G. Tzanetakis, and P. Cook. "Musical genre classification of audio signals." *IEEE Transactions on speech and audio processing,* 2002.
[534] G. Peeters "Template-based estimation of time-varying tempo". EURASIP JASP, 2007.

based on deep learning. These models are automatically learned using a dataset of annotated song recordings, for which the tempo is known[535].

These technology advances have a positive impact on the analysis and the processing of the music at the ***signal level***, but also at the ***symbolic level***, that is the score of a piece of music. In the context of the creation of contemporary and experimental music, some computer tools have been developed to help musicians when composing and orchestrating innovative music pieces[536]. Inspired by serialism[537] and spectralism music[538], some approaches are able to generate musical excerpts based on chosen constraints or a given musical sample[539] (Figure 104).

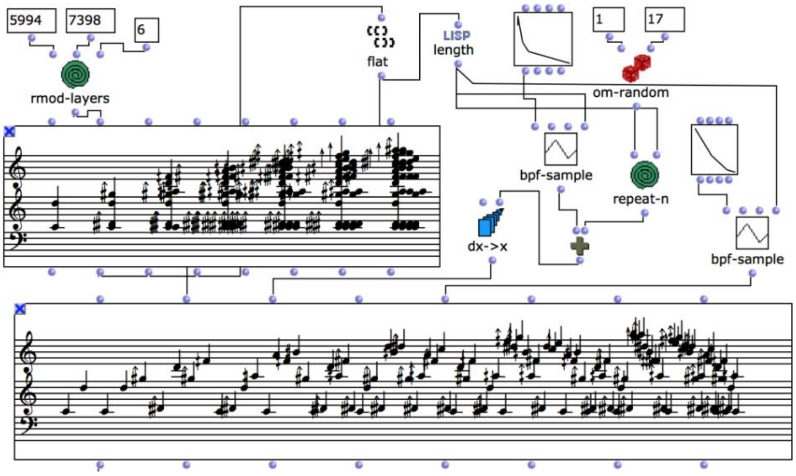


Figure 104: Automatic music generation tool based on rules and constraints[540].

Previously limited to experimental music, for which it is often possible to formalise the musical concept by mathematical constraints, nowadays this research field tends more towards ***popular music***, thanks to deep learning, among other reasons. Indeed, compared to classical and experimental music for example, popular music is more related to the musician's feeling, quite difficult to describe, rather than intellectual concepts. So, based on relevant and numerous examples, the flexibility of the deep neural network models makes possible to infer more sensitive concepts in music.

Research challenges

535 H. Foroughmand, and G. Peeters, "Deep-rhythm for tempo estimation and rhythm pattern recognition", ISMIR, 2019.

536 J. Bresson, C. Agon, and G. Assayag. "OpenMusic: visual programming environment for music composition, analysis and research." *19th ACM International Conference on Multimedia*. 2011.

537 Wikipedia article on Serialism: https://en.wikipedia.org/wiki/Serialism

538 Wikipedia article on Spectral music: https://en.wikipedia.org/wiki/Spectral_music

539 C. Cella, et al. "OrchideaSOL: a dataset of extended instrumental techniques for computer-aided orchestration." 2020, arXiv preprint arXiv:2007.00763.

540 Open Music software, image source: http://christophertrapani.com/wordpresssite/computer-assisted-composition/

One standard problem in machine learning, and especially in deep learning, is the ***lack of data***. For example, traditional tempo estimation models are learned using a supervised training which needs some songs labelled to their tempo. First, the size of the training dataset must be as large as possible; second, the used examples must be of a good quality with a number of wrong annotations as low as possible; and third, the used dataset must be representative of the world of interest. For example, a tempo model trained with classical music may fail with electronic songs.

Unfortunately, for most of the applications in music processing, the creation of such datasets is difficult and can be expensive. To deal with this problem, many research results have been proposed for machine learning, and more specifically deep learning: data augmentation, non-supervised generative models (e.g. Variational Auto-Encoder), student teacher learning, data synthesis, domain adaptation, generative adversarial networks, differentiable processing, and transfer learning. This data issue is not restricted to music and sound processing, it is also a strong limitation for most of the other applications of deep learning. Even if it is possible to mix these mentioned methods, a challenge to researchers is to continue to develop ***new learning methods able to be trained with scarce data***.

A second challenge for AI research in music and sound processing is the "***artistic innovation***". In music history, the evolution has been usually preceded by technological advances. For example the romanticism of the 19th century is partly linked to the creation of the piano, which replaced the harpsichord of the 18th century. Nowadays, the use of AI can provide new ways to create music, and one could think that it makes possible the emergence of new styles. Nevertheless, there is the risk of the opposite effect. Indeed, because deep learning models are trained on existing music examples, it seems impossible for a machine to properly imagine the music of tomorrow; ***humans must stay in the loop***.

Figure 105: Graphical Interface for a smart exploration of a catalogue of drum sounds[541].

[541] Software Xo of xln-Audio, image source: https://www.xlnaudio.com/products/xo

Consequently, most of the adopted strategies consist in providing an ***artificial assistant*** to musicians, but the question of creativity remains. For example, among the current AI assistants, the most advanced tools propose some generated musical samples according to the inputs given by the musician (Figure 105). Sometimes he/she has the possibility to transform the solution by moving in a low-dimensional latent space. For now, these algorithms provide limited degrees of liberty to the musicians, and a challenge for next research in music and sound processing is to ***give more control to the musicians*** to enhance the musical innovation. In other words, the musicians should be able to explore sounds and music out of the current distributions, but in an artistic and explainable way, and not in a random way.

Societal and media industry drivers

Vignette 1: Composing music for a movie soundtrack with the help of an AI assistant

Joannah is a composer for a movie soundtrack. Her friend Stevia is a movie director, and she commissioned the composition and the orchestration of her next movie. During the shoot, Joannah composed the musical themes. Now the video editing is finished and she needs to align the orchestration on the video before the rehearsal and the recording with the musicians. Unfortunately the time is short, so she uses her AI assistant for the orchestration and the synchronisation of the music and the video. This consists of modifying the prepared music score to align it in time with the fixed video according to the ambiances, e.g. romantic, suspense, action, and also to punctual events. Joannah does not want the AI assistant to compose for her, she wants an intelligent assistant able to help her during the orchestration but without Joannah losing the control on the generated score. Thanks to it, Joannah prepared the music on time, and using the integrated sound synthesiser, she could produce a first preview, which Stevia validated before the recording with the real musicians.

Vignette 2: AI assistant for automatic music arrangement and sound mixing in live performances

Tom and Guillem form an electronic music band which is popular in the nightclubs. They like to improvise in live and to experiment with new things every night. Rather than to mix different sound recordings or drum machine samples, they use an AI assistant, which is able to generate different music tracks (e.g. bass lines, drums, synth drones) automatically adapted to the sound features of the played music, e.g. rhythm, tempo, harmony, ambiance. Contrary to standard drum machines, this assistant knows their particular musical style that differentiates them from other DJs. It is a great help to manage several instruments during the improvisations; for example, Tom and Guillem can decide to suddenly change the ambiance, and the assistant automatically changes the music tracks accordingly to the parameters controlled by the duo. They like it because, they keep full control on the assistant, it frees them from having to do boring tasks, it fits their musical style, and it improves their creativity during live performances.

Future trends for the media sector

The analysis and the generation of music already has an important role in the media sector. Nevertheless, there is a room for AI to improve its impact on media. Some future research trends are summarised below:

- Improve ***musical recommendation systems***, e.g. for video games, radios, documentaries, a) to better fit the user taste and the targeted features, and b) to take into account the context of the listening.
- Allow ***training on small datasets*** created by the end users for a specific task: a) improve the learning ability with scarce data, and b) design generalised models able to learn different tasks without the need to be used by expert users.
- Improve the ***learning of emotions*** provided by the music and felt by humans: a) for the analysis of music recordings, and b) for the generation or treatment of musical sounds and scores.
- Improve the ***learning of the musical language***. This could also improve the analysis and the synthesis of musical audio signal.
- Give ***more controls to the musicians*** during the process. For example, propose some understandable and controllable constraints which makes sense from an artistic point of view.

Goals for next 10 or 20 years

Musical AI applications will provide an improved understanding about what makes sense in the music: style, emotion, etc. But the machine should not be a competitor for musicians; it should stay an assistant which will be able to help them when doing tedious, repetitive, or long tasks. It is illusory to think that an AI model can be innovative in music and can propose new styles in a proper and artistic way, without a random process. For music creation, humans must stay in the loop, in consequence the research will be active to find some ways to easily control the AI process.

Nowadays, most AI applications for automatic composition and arrangement are based on digital scores in the MIDI format. First, we can have access to thousands of MIDI files but with a poor quality, without realistic nuances and variations; secondly, quality digital scores are few; and third, pairs of sound mixes with quality scores are even rarer. Nevertheless, we can access millions of recorded songs with high audio quality, many of them having partial annotations, but without the associated scores.

The main goal for the next 5-10 years will be to make the link between scores and sound mixes. To achieve this goal, some advances in deep learning are particularly interesting: for example because we do not have synchronised audio-score data, unsupervised methods and generative adversarial networks are very promising approaches, and some interesting research projects have already started. For a longer-term perspective, in the next 10-20 next years, using this future knowledge on the relation between symbolic music and sound recording, we will be able to learn efficient embedded musical representations of compositions and performances. These models will implicitly learn all the standard musical rules but also how to break them in an appropriate way like many talented musicians do.

Finally, a strong limitation still remains in capturing the emotions and moods provided by a song. A future solution can be given by neuroscientists. Indeed, the study of emotions is an active topic in neuroscience and we can expect that future results can be reused in musical AI applications. For example, we can imagine that in two decades, the emotion of a human subject will be efficiently detected by lightweight and portable encephalographic devices, and that volunteers will help in making emotion datasets just by listening to the music when travelling by train or by plane.

9.9 AI for publishing: from customer intelligence to prosumer intelligence

***Contributors**: Julià Vicens, Mihnea Tufis, Cristian Consonni, Patricia Castillo, David Laniado (Eurecat - Centre Tecnològic de Catalunya)*

The authors of this contribution are part of the **Möbius *H2020 research project***[542]*, which aims to modernise the European book publishing industry by remodelling the traditional value chains and business models uncovering the prosumers potential and delivering new enriched media experiences.*

Current status

Vendors of products from very diverse sectors have been implementing increasingly sophisticated processes for capturing and analysing data from their customers in order to understand their specific needs, to enrich the catalogue of products, and to iteratively improve the individual relationship with each customer. These processes, denoted customer intelligence, have been favoured precisely by the digitisation of products and interactions with customers. Nevertheless, the consumer is considered as a passive entity, without accounting for their creative potential. The traditional frontiers between vendors and customers in media markets are blurring. Therefore, the times call for a better understanding of the prosumer phenomenon as a key aspect for business competitiveness.

Although user-centric innovation began to be discussed in the 1970's, it is not until very recently that evidence on the impact of involving users in innovation processes for improving business strategies started to arise[543]. The concept of "prosumer" is not new, either. A decade later, in

[542] Möbius project (funded by H2020 under grant agreement no 957185): https://mobius-project.eu/
[543] Gamble, J. R., McAdam, R., & Brennan, M. (2019). How User-centric Innovation is Affecting Stakeholder Marketing Strategies: Exploratory Findings from the Music Industry. *European Management Review*. https://pure.ulster.ac.uk/ws/portalfiles/portal/12590625/Gamble_et_al._2018_.pdf

1980, Alvin Toffler coined the term to signify the people who produce many of their own goods and services, and will tend to replace consumers in the transition from the Industrial to the post-industrial Age.[544] Izvercian et al (2013)[545] define prosumers as usual consumers "*eager to engage, malleable to be transformed and creative enough to produce valuable outcomes*". Importantly, they "*become empowered and are aware of their worth*". Thus, they have potential for collaboration and co-creation of value that may be useful for corporate interests across various industries and markets.

Wattpad online social reading platform is already capitalising on its overwhelming amount of original stories created by their users, and the data they generate in the process. For Wattpad, an open community of writers and readers, prosumers become sources of fresh ideas.[546] The successful Netflix original film "The Kissing Booth" is based on a 2011 story published in Wattpad by a 15-year old that was read by 19 million people on Wattpad before it was turned into a series of books. To keep on this track, Wattpad is partnering with media networks and producers to adapt Wattpad stories for television and has recently launched a Book Division that levers on their Story DNA machine learning technologies for combining data driven approaches "with human editors' critical eyes".[547]

Another iconic example illustrating the power of prosumer communities is the publishing phenomenon "Fifty Shades of Grey" by author E.L. James. The trilogy was initially a fanfiction creation based on the successful Twilight series (a vampire-themed fantasy romance). After some reworking and interactions with fellow fanfiction followers, James self-published – in e-book and print-on-demand formats – the first installment of her trilogy in 2011 through Australian virtual publisher "The Writer's Coffee Shop." Viral marketing did the rest and "Fifty Shades of Grey" managed to sell 100 million copies worldwide and had its first movie premiered in 2015, only four years after the book was first published.[548]

While current practices of publishers are still mostly based on a vision of the consumer as a passive actor who will just buy or not buy the product, and more or less traditional marketing and recommendation approaches are followed in order to maximise sales, the potential of the prosumers as co-creators of content and in all the steps of the process is not fully taken into account, and the enormous wealth of content and interactions generated by these online communities remains still untapped (Figure 106).

Furthermore, popular fanfiction platforms like Archive of Our Own (AO3) or fanfiction.net, where thousands of users co-create content every day, are based on relatively old technology

---

544 Toffler, A., & Alvin, T. (1980). The third wave (Vol. 484). New York: Bantam books.
545 Izvercian, M., Seran, S. A., & Buciuman, C. F. (2013). Transforming usual consumers into prosumers with the help of intellectual capital collaboration for innovation. *International Journal of Information and Education Technology*, 3(3), 388. http://www.ijiet.org/papers/304-N00018.pdf
546 C. Lee, How Wattpad Is Rewriting the Rules of Hollywood (2018): https://www.vulture.com/2018/07/how-wattpad-is-rewriting-the-rules-of-hollywood.html
547 C. de León, Wattpad, the Storytelling App, Will Launch a Publishing Division (2019): https://www.nytimes.com/2019/01/24/books/wattpad-books-publishing-division.html
548 Statista, Number of copies E.L. James's 'Fifty Shades of Grey' sold in selected countries worldwide as of February 2014: https://www.statista.com/statistics/299137/fifty-shades-of-grey-number-of-copies-sold/

and offer a simple interface that does not take advantage of the advances in UX research from recent years, or on automated algorithms even for established tasks such as filtering and recommending content to the users. There is room for improving user experience and collaboration practices in this field, with AI at the service of the users and the communities.

In the Möbius project[542], focused on "The power of prosumers in publishing," funded by the European Commission under the Horizon 2020 programme, the aim is to develop methods and tools to effectively streamline cooperation with prosumers – including from open communities – in publishing workflows, and in particular: to leverage on authors and influencers for informing content innovation and publishing decisions and establish cooperation strategies; and on beta-testers and early-adopters for innovation processes for new products or experiences.

Figure 106: In fanfiction communities, prosumers build on existing work of fiction to write new stories in a canonical fictional universe, such as the one from Harry Potter, represented in the figure. Photo by Jules Marvin Eguilos on Unsplash[549].

Research challenges

In the publishing sector, taking advantage of the wealth of data created by prosumers, in the form of original content, reviews, feedback and interactions, implies being able to mine this data, make sense of it and extract actionable knowledge for improving the community dynamics, productivity and synergies, and for ultimately publishing and promoting successful and innovative products.

While a big effort has already been spent in previous research to develop methods and algorithms for mining social media and online platforms, some particularities of the context of fanfiction and prosumer communities make it unique, so that it needs to be addressed by a specific modelling effort.

This includes identifying and characterising different prosumer profiles, according to the capacity in which they are making a contribution. As a first classification, we can see:

- *Authors* create new stories, even if they are fandom contents.

[549] Image source: Jules Marvin Eguilos, Unsplash, https://unsplash.com/photos/YrcfSDVli3Y

- *Influencers* may also be generating original content, although not to create new stories, but to generate discussion about them; and so they are able to engage communities and elicit contributions around this shared interest.
- *Beta-testers* take part in the development of new products and services, their feedback is crucial for ensuring the innovation meets user expectations, and that it is meaningful, useful and acceptable.
- *Early-adopters* are the groups most likely to adopt or consume an innovation as soon as it is released, so their reaction to the near-to-final product is of utmost interest. Importantly, they all bring user knowledge, interactions, opinions, and help generate further user requirements that would typically elicit new user-driven innovation processes.

Identifying different profiles, roles, and levels of involvement and reputation of prosumers is crucial for being able to engage specific users for specific goals and tasks. Furthermore, it is part of mining community dynamics in order to monitor the state and health of the community, and to be able to take action to improve it.

Collaboration and co-creation happen spontaneously in prosumer communities, however the interaction mechanisms allowed by the platform play an important role: they may foster more or less fruitful cooperation dynamics, encourage synergies and distributed mentorship,[550] or in some cases on the contrary they may even hinder collaboration, or incentivise toxic behaviour. AI in this sense can help experiment and find out what works better, not only with empirical analysis of digital traces from existing communities, but also with practices such as A/B testing under tight ethical codes.[551]

Knowledge extracted from the interactions and community dynamics can be complemented by mining the textual content of the works produced by the users, extracting from the text features such as topics, emotions, characters and entities, and linguistic styles, quality, and diversity. Combining textual features extracted from the content with interaction features and social dynamics around the corresponding content may open up a great potential for innovation, where AI can help make the co-creation process more effective and satisfactory for prosumers.

We believe in this context that not only a human-centric approach[552] is fundamental, ensuring algorithmic transparency and explainability, and keeping a watchful view on bias and discrimination potentially introduced by the systems, so that AI is actually at the service of the user, who is able to make sense of the algorithmic decisions; but also a community-centric

[550] Campbell, J., Aragon, C., Davis, K., Evans, S., Evans, A., & Randall, D. (2016). Thousands of positive reviews: Distributed mentoring in online fan communities. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 691-704).
[551] Benbunan-Fich, R. (2017). The ethics of online research with unsuspecting users: From A/B testing to C/D experimentation. *Research Ethics*, 13(3-4), 200-218.
[552] Xu, W. (2019). Toward human-centered AI: a perspective from human-computer interaction. *interactions*, *26*(4), 42-46.

approach, "keeping community in the loop" as proposed in recent research on the Wikipedia community, to make sure algorithms are aligned with the values and needs of the community.[553]

Societal and media industry drivers

Vignette: Publishing a manuscript from an online fanfiction community

Ginette works at a publishing company, and her duty is to select manuscripts for publication. Ginette does not only examine manuscripts she receives, but she works in collaboration with the online community around a co-creation platform, and relies on an AI monitoring system that discovers promising content.

Among the hundreds of works that are created every day on the platform by the users, the system points Ginette to a manuscript that has been created just a few hours earlier, and has already received positive feedback from an influential trend-setter from the community.

Based on the very first feedback, and on the linguistic, topical and emotional features automatically extracted from mining the manuscript's text, the system has selected this as a promising manuscript, and shows Ginette a bunch of synthetic indicators estimating its linguistic style, quality, emotional content, main topics, main characters from the original book from which it is inspired, relation to other manuscripts published by the company, potential target audience in terms of demographic characteristics and other marketing criteria.

With a look at the indicators, Ginette immediately sees an opportunity in this manuscript, and activates the "early feedback" program. The system identifies the users who have the best match with the features of the story, and at the same time have a relevant experience and position in the network of users of the platform, and invites them to review the manuscript. So, beyond the spontaneous reviewing process that will happen within the community, these users are pro-actively invited to give their feedback. The users will have the possibility to give suggestions to the author, and eventually become co-authors.

Once various rounds of reviews are completed, the book enters the final phase of the preparation for publication by the publishing company. The book that is launched to the market is in a way a product of the whole community that is involved also in creating images and messages for promoting it in social media.

Future trends for the media sector

AI may help to develop prosumer intelligence for the publishing sector, by providing algorithms and techniques for taking advantage of the untapped potential in the wealth of data generated by the users, and leveraging it for improving the co-creation and dissemination processes. This can be achieved addressing several issues and goals:

[553] Smith, C. E., Yu, B., Srivastava, A., Halfaker, A., Terveen, L., & Zhu, H. (2020, April). Keeping community in the loop: Understanding wikipedia stakeholder values for machine learning-based systems. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-14).

- *Early detection of trends in content consumption*: Identify content and topics that will become popular in the near future, and that have potential for publishing; with algorithms based on temporal patterns, interactions, and features of user generated content.
- *Early detection of trends in content production*: identify topics, styles, emotions, genres and subgenres, patterns that are increasingly adopted by authors, with particular attention to users that tend to establish or adopt early new trends that will spread in the community.
- *Fostering synergies and distributed mentoring*: identify users that could fruitfully collaborate with each other or review one another's work, based on common interests, and complementary abilities, expertise and styles.
- *Identification of relevant users*: mine user behaviour to identify prosumers that assume specific roles in the community: authors, fans, influencers, beta-testers, early-adopters, trend-setters.
- *Selection of seed content*: identify content and stories, from the community itself as well as from external sources, which can be used as input to stimulate the co-production of new content.
- *Monitoring of community dynamics*: identify phases and trends in the growth of a prosumer community, and actions needed to improve community health.
- *Annotation mining*: enrich the reader's experience not only based on feedback, comments and annotations from other users, but also on further predictions based on such user generated content.
- *Indexing of content*: automatic identification and indexing of topics, characters, objects and other kinds of entities, in order to allow for browsing and exploring content, and for connecting works and manuscripts with each other at different levels of granularity, including by chapter or paragraph.

Goals for next 10 or 20 years

When we consider prosumer communities devoted to co-creation of content, such as fanfiction communities, we have to be aware that at the very core of these socio-technical systems we have humans, both at the individual level with their creativity beyond their personal tastes, preferences and habits, and at the community level with their interactions, their social dynamics, their collaboration practices, and their belonging to a social group with shared values and objectives. Therefore, we believe the most crucial goals for AI in this field have to do with being able to put algorithms at the service of users and communities, integrating into their daily practices and aligning with their needs and values. AI algorithms can act as a multiplier of inequality and reinforce existing bias, therefore they must be designed keeping in mind their effect on the people who use them or who may be affected, which must be able to participate and control how the AI is used. We envision a shift from currently dominant black-box models and opaque algorithms run by the platforms to maximise variables such as clicks or time spent on the platform, to an emerging paradigm of human-centric and community-centric AI at the service of the communities, with transparent algorithms that are able to foster social connections, synergies and human creativity.

9.10 AI for social sciences and humanities research

***Contributors**: Rasa Bocyte (NISV), Philo van Kemenade (NISV), Johan Oomen (NISV)*

Current status

The emergence of ***digital humanities*** (DH) has fundamentally changed the way ***social sciences and humanities*** (SSH) researchers can perform research on media content. The ***Computational Turn*** allows scientists to ask completely new research questions and investigate existing research questions in new ways.[554] While previously the field was dominated by "close reading" methods, in which individual sources are analysed, DH brought about a turn towards "distant reading" that allows researchers to scrutinise and compare a large number of sources, spanning across extended periods of time, languages and media formats. For instance, researchers now can investigate gender representation in the media by comparing air time given to male and female speakers on television, and analyse how this has changed throughout years. With the help of AI-based research methods, it became possible to detect and observe patterns from an extensive historical perspective (Figure 107).

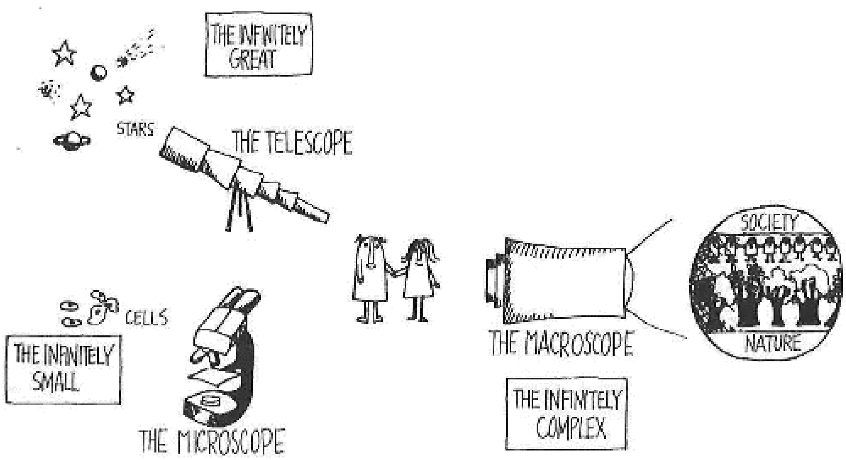


Figure 107: Microscope / Telescope / Macroscope[555].

In Europe, these developments in the SSH field would not have been possible without significant investments made over the last decade in the heritage sector, both towards digitising archival materials and also ensuring that born-digital materials (ranging from documents to social media

[554] S. Wyatt. "A Computational Turn in the Humanities? A Perspective from Science and Technology studies." *Journal of Siberian Federal University* 9, 2016, 517-524.
[555] Drawing by Joël de Rosnay: https://scottbot.net/the-historians-macroscope/

data) are archived. Recognising the research potential of these digital materials, several pan-European initiatives such as CLARIN[556], DARIAH[557] and E-RIHS[558] have emerged to support the SSH field. These initiatives, supported by universities and cultural heritage organisations, are developing infrastructures and tools that enable DH research as well as building the capacity of researchers to work with computer-based research methods (Figure 108). Alongside them, pan-European data aggregators and providers such as Europeana[559] provide access to carefully collected and curated multimodal data that have huge potential for training machine learning models.

From the outset, DH scholarship was largely dominated by text-based analysis.[560] This is due to the early maturity of AI-based tools for text recognition, as well as the availability and accessibility of large corpora of textual content as training data. With the growing maturity of computer vision techniques, availability of processing power, analysis of visual culture is increasingly gaining prominence, which is an essential development for SSH scholars who work with multimodal data.[561]

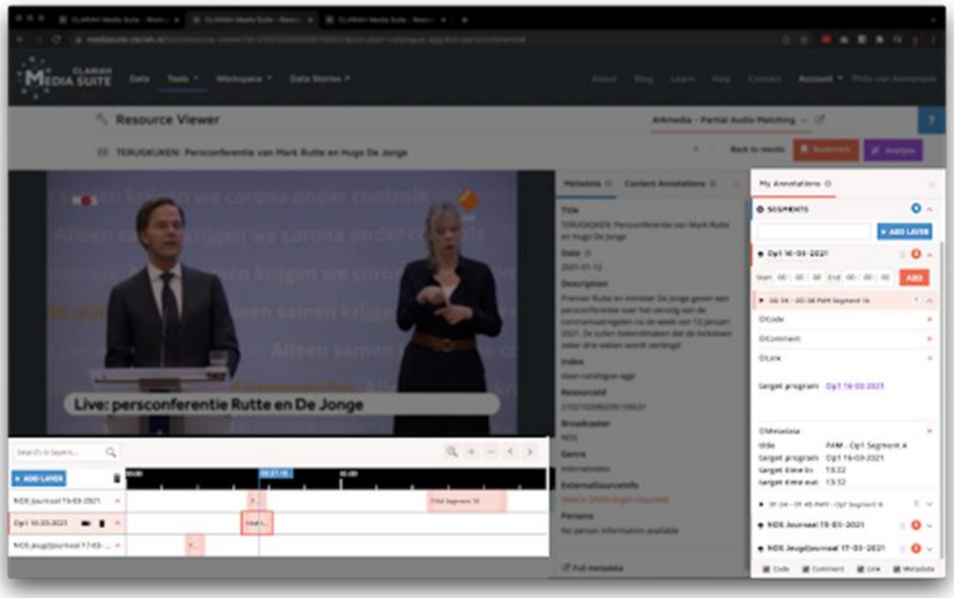


Figure 108: Interface of the CLARIAH Media Suite[562] research infrastructure for SSH scholars showcasing partial audio matching functionality that enables researchers to track the reuse of a particular audio segment in other programmes.

556 CLARIN (Common Language Resources and Technology Infrastructure): https://www.clarin.eu/
557 DARIAH-EU (DIgital Research Infrastructure for the Arts and Humanities): https://www.dariah.eu/
558 E-RIHS (European Research Infrastructure for Heritage Science): http://www.e-rihs.eu/
559 Europeana: https://pro.europeana.eu/
560 M. Wevers, T. Smits, The visual digital turn: Using neural networks to study historical images, *Digital Scholarship in the Humanities*, 35:1, 2020, 194–207, https://doi.org/10.1093/llc/fqy085
561 T. Arnold, S. Scagliola, L. Tilton, and J.Van Gorp, Introduction: Special Issue on AudioVisual Data in DH. *Digital Humanities Quarterly. Special Issue: AudioVisual in the Digital Humanities*, 2020. http://www.digitalhumanities.org/dhq/vol/15/1/000541/000541.html
562 CLARIAH Media Suite: https://mediasuite.clariah.nl/

Research challenges

A number of challenges need to be addressed to fully exploit the potential of AI-based research methods in SSH:

Processing big multimedia data. Media collections are massive and the individual items are rich with information, which means that performing even basic analyses requires large amounts of storage and processing power. From the perspective of a researcher, this means significantly limiting the scale of their investigation as performing research on the entire media collection held by an organisation might not be attainable. From the perspective of an organisation facilitating access to their collections via AI-based tools, the processing power required leads to significant costs, which means that not all organisations can afford this. This scalability challenge of processing big data is particularly pertinent in broadcast archives, where the collection grows every day with industry-standard high-quality audiovisual content. The negative environmental impact of such high-demand AI systems is also a great concern that calls for solutions for optimising algorithms and infrastructure for energy efficiency.

Managing the complexity of cultural heritage data. Deep-learning algorithms may yield impressive results in supervised training on data when given large numbers of labelled patterns but its utility decreases in the context of digital collections such as news content where the continuous stream of new annotations and changing interpretations of existing data hamper a better understanding of multimodal meaning making.[563] The data-hungry nature of deep learning algorithms obstructs their use in sparse or incomplete data situations such as those in the cultural heritage domain. Moreover, heritage collections pose an epistemological challenge to AI systems as a multitude of different, even conflicting meanings that coexist and evolve over time can be associated with a single cultural object, making it difficult to classify them.

Usability of AI systems for SSH researchers. Due to the complexity of AI systems, the availability of user-friendly AI tools for media scholars is very limited. The entry barrier for a researcher to start using such tools is high and often requires additional training. There is a lack of interdisciplinary understanding between those who are developing the tools (AI experts) and those who would benefit from them (SSH researchers). Equally, SSH scholars do not receive sufficient digital training to (i) benefit from the opportunities opened by AI tools, and (ii) execute tool criticism.[564] While commercially available tools often provide more easily-accessible interfaces, they are too generic to support the complex analyses required by the SSH scholars. Media content holders implementing AI tools in their technical infrastructures face similar skill-related challenges. AI skills especially in the cultural heritage sector are rare, making it difficult for such organisations to support SSH researchers with cutting-edge AI technologies.

[563] Muccini, H. & Vaidhyanathan, K. (2020). Towards self-learnable software architectures, ERCIM News 122, p. 33-34.

[564] M. Koolen, J. van Gorp, J. van Ossenbruggen, Toward a model for digital tool criticism: Reflection as integrative practice, *Digital Scholarship in the Humanities*, Volume 34, Issue 2, June 2019, Pages 368–385, https://doi.org/10.1093/llc/fqy048

Lack of training data. IP regulations built on protecting the commercial interests of media content creators limit the accessibility of professionally-produced audiovisual content available for model training. Due to this, many available AI systems are trained on datasets from widely accessible online platforms such as YouTube, that have not undergone rigorous quality checks and considerations for biases that such datasets would introduce. Moreover, people whose data have been included in such datasets have not been informed about this and have not given their consent. Often, this results in algorithms that are prejudiced against non-Western societies and minorities.[565] Researchers working with such AI systems might not even be aware of the provenance of the data used to train models and what repercussions it could have on the results as the biases might not be immediately visible. While the discourse around the need for more diverse training data is gaining momentum in the academy and the commercial sector, there is still a lack of standardisation and best practices for creating and documenting datasets.

Generic AI models. Currently used off-the-shelf AI models are not suitable to address the complex research of SSH scholars. In particular, they are lacking in understanding of cultural and historic sensitivities, domain-specific knowledge, co-existence of multiple meanings and interpretations defined by the context.[566] Additionally, translating purely visual or abstract concepts that are relevant for SSH is a complex undertaking that requires high-quality annotations from topical experts. Techniques that enable adjusting or "fine tuning" of pretrained generic models to a particular context (such as transfer learning) are promising in this regard.

Technology-driven development. Many of the above-listed challenges are to some extent caused by the lack of interdisciplinary approaches applied to the development of AI. All too often, AI engineers and domain experts speak completely different languages which prevents meaningful collaboration. In most projects, the intersection between technology and SSH is not fully realised (e.g. humanities researchers are involved only to provide the initial user needs and test the final outcomes, but they are not involved in the decisions regarding model training). Hence, in many AI initiatives from academia considerations for biases and user-friendliness come as an afterthought rather than primary requirements.

Societal and media industry drivers

Vignette: AI-enabled Social Sciences and Humanities research

Luna is a psychology professor leading a lab of researchers who study cultural differences in body language. She is collaborating with an audiovisual archive that provides access to its collection via a research tool with integrated AI solutions, including identification of gestures and poses which are essential for Luna's research. Before using the tool, she consults documentation about the algorithm behind the tool - this information is displayed via non-technical visualisations that allow her to immediately see what data was used for training. She

[565] J. Buolamwini and T. Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, *Conference on Fairness, Accountability and Transparency*, 2018, 77–91.
[566] T. Smits and M. Wevers, "The Agency of Computer Vision Models as Optical Instruments," Visual Communication, 2021, https://doi.org/10.1177/1470357221992097

quickly notices that training datasets predominantly consist of content from contemporary Western data sources which would skew her research results. She also sees the algorithm's provenance - it was taken from a European online platform of open-source pre-trained AI models for media archives. On the platform, she starts a discussion about the concerns she has and quickly finds a critical mass of scholars who are also interested in addressing this. Led by Luna's team, together they create a small dataset with high-quality annotations, mainly using historic content that the archive provides to them. This dataset is then used to fine-tune and improve the AI model and is integrated into the archive's research infrastructure. Luna's team also documents decisions made when creating their dataset and annotations, and publishes it on the online platform for other scholars to consult and improve in the future.

Future trends for the media sector

With the above-listed challenges in mind, we foresee the following trends that would both strengthen the use of AI for SSH scholarship and support AI research through SSH involvement.

Trends related to technological advances

High-quality annotations. While model training requires high volumes of data rendering many quality control processes unscalable, even small amounts of high-quality annotations can significantly improve models. SSH scholars can play an important role in providing domain-specific annotation and meaning, which can be used to increase the usability of AI tools for research purposes as well as be transferred to other application domains. Here it is important to recognise that just like textual data, visual language is also culturally-specific therefore including media scholars with pluralistic perspectives in annotation efforts is key to avoiding skewed results. Equally, heritage organisations should increase their efforts on treating their collections as potential training datasets. This should inform their data enrichment strategies – for instance, consulting communities represented in the data to ensure that their perspectives are adequately captured and represented fairly.

Focus on interface design. User-friendly interfaces that do not require advanced computational skills will play a key role in boosting the uptake of AI-based research methods by SSH scholars. However, usability should not come at the cost of reduced complexity. On the contrary, developments in this domain should be driven by rigorous research methodologies that SSH researchers require, prioritising transparency and explainability.

New media formats. New media formats such as web-based, VR and XR productions are increasingly entering the media sector and archival collections. Analysis and creation of annotations for these non-linear, multilayered media artefacts ask for new solutions. While this challenge does not stem directly from the SSH field, we strongly believe that SSH scholars, especially (new) media experts, can play a key role in these research efforts.

Trends related to Ethical Legal and Societal Aspects (ELSA)

Opening black boxes. Users know little about the ways in which their queries produce results. Currently used algorithms, while expediting access, usually do not account for potential bias, incompleteness of data, unless they are designed and trained to do so. New generations of AI solutions need to remedy this lack of knowledge by developing self-correcting, contextualising and human decision-enabling tools that increase the user's confidence and ability to "make sense" of the data.

Shared infrastructures. Shared infrastructures that support the exchange and reusability of datasets, trained models and best practices will provide a significant push towards the emergence of better tailored AI solutions for SSH and the wider media sector. In particular, the creation of Data Spaces will offer opportunities to train algorithms on more diverse, cross-institutional datasets.

Convergence of models. AI models that can perform multiple tasks dealing with different modalities of media content (combining visual, textual and audio analysis) will render AI-based research methods much better suited for the needs of SSH scholars.

Interdisciplinary projects. The growing number of SSH researchers equipped with computational skills is going to result in more meaningful interdisciplinary collaboration in the field. Currently, many projects at the intersection of AI and SSH are still driven by either technology or theory. In the future, we hope to see theoretical approaches from SSH playing an equally important role as technology-focused research, and in particular strongly permeating data collection and model training practices.

Goals for next 10 or 20 years

Towards customisability. SSH scholars approach digital tools with requirements that are very specific to their research questions and cannot be always satisfied with generic solutions. The attractiveness of AI techniques in this field will dramatically increase with the growing customisability of AI-based research tools. Users should be able to adjust and experiment with the parameters of such tools and fine-tune existing models with custom concepts relevant for their research (for instance, by using few shot learning), as well as set up collaborative experimentation environments where they can compare their analyses. This would not only enable researchers to get more satisfactory and meaningful results, but also improve their overall trust in AI techniques and critically engage with them.

Aim for meaningful interdisciplinarity. With a growing number of concerns about biases and social injustices replicated and amplified by commercial AI systems, the intersection between AI experts and SSH scholars has the potential to grow in significance, with the goal to question current practices and collaboratively develop more equitable solutions. The critical analytical approach that SSH scholars apply when working with AI tools would result not only in better tailored research tools but also produce better AI models and practices that could be transferred to wider societal contexts. To achieve this, it is of essence to strengthen interdisciplinary education efforts. While digital literacy skills are becoming more commonly introduced in the humanities, the same cannot be said about exact sciences - AI and machine learning curricula lack the humanities perspective that is essential to redress algorithmic biases. Design could act as an intermediary discipline that bridges the two sides together, making it easier for technologists to translate user needs into technical requirements and interface design features and for humanities scholars to interact with and question AI tools.

Towards more equitable AI research. Access to high-quality big data will be one of the determining factors of Europe's leadership in AI for SSH and the media sector at large. Currently, social media companies dictate the rules in the market by determining who gets to access their vast datasets necessary for model training. In the next ten years, we hope to see heritage organisations emerging as strong competitors in this domain, offering access to high-quality, culturally-aware and contextualised datasets. To get there, we need to see concerted advocacy efforts from the European media industry and the research community for radically increasing the openness of media collections, ensuring that scholars and machine learning engineers have the right resources and skills to develop AI tools. The European legislative framework for AI[567] can play a crucial role here by positioning open collaboration and resource sharing as essential ingredients of AI research that can elevate Europe's economic and societal wellbeing.

567 European Commission, A European approach to artificial intelligence: https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence

10 Open AI repositories and integrated intelligence: next steps towards AI democratisation

***Contributors**: Sven Becker (FhG-IAIS), Annerike Flint (FhG-IAIS), Laszlo Friedmann (FhG-IAIS), Andreas Steenpass (FhG-IAIS)*

In this section, we discuss issues related to AI democratisation, focusing on open repositories for AI algorithms and data and research in the direction of integrated intelligence, i.e. AI modules that will be easily integrated in other applications to provide AI-enabled functionalities.

Current status

The media sector is already applying AI technologies. For example, users consuming media already enjoy live subtitling, automatic translation into other languages, or into plain language. One might argue that these mechanisms have introduced ***integrated intelligence***, i.e. adding a software component that makes a product intelligent, such as live subtitling for a live video stream.

To design an AI system, one needs data, storage and computational resources, algorithms, and AI talent. On all these levels, ***openness*** plays an important role and stretches across multiple levels of AI development.

When it comes to data, thousands of open source datasets are published on different platforms like GitHub (awesome-public-datasets[568]), Kaggle[569] or by Google[570], to name a few. The limitation of storage and computational resources was addressed during the last years by three major AI cloud providers: Amazon Web Services (AWS), Microsoft Azure and Google Compute Platform. In November 2018 GitHub announced that 100 million code repositories are live on GitHub. According to Jason Warner, the repositories stem from 31 million developers from almost all countries in the world, which are collaborating across 1.1 billion contributions.[571] In recent years Microsoft, Google and Amazon have provided development tools to allow practitioners to create AI applications without deep knowledge of machine learning, linear algebra, or statistics. This addresses the issue of limited AI talent and the transfer of AI tools into multiple industry sectors. The provided technologies are designed as ***modular tool boxes***.

In general, the widespread use of open source software has increased in popularity. TensorFlow, a machine learning framework for basic machine learning model development (developed by Google), is in the top 10 of all GitHub projects with over 150,000 star ratings.[572] However, the importance of open source is also reflected in the availability of much more specific AI

[568] Awesome public datasets: https://github.com/awesomedata/awesome-public-datasets
[569] Kaggle datasets: https://www.kaggle.com/datasets
[570] Google public datasets: https://cloud.google.com/bigquery/public-data/
[571] Warner, Jason. "Thank you for 100 million repositories". Last accessed: 16 December 2021.
[572] GitHub Ranking, top 100 stars:https://github.com/EvanLi/Github-Ranking/blob/master/Top100/Top-100-stars.md

frameworks for solving tasks for example in the domain of Natural Language Processing (NLP) (Figure 109).

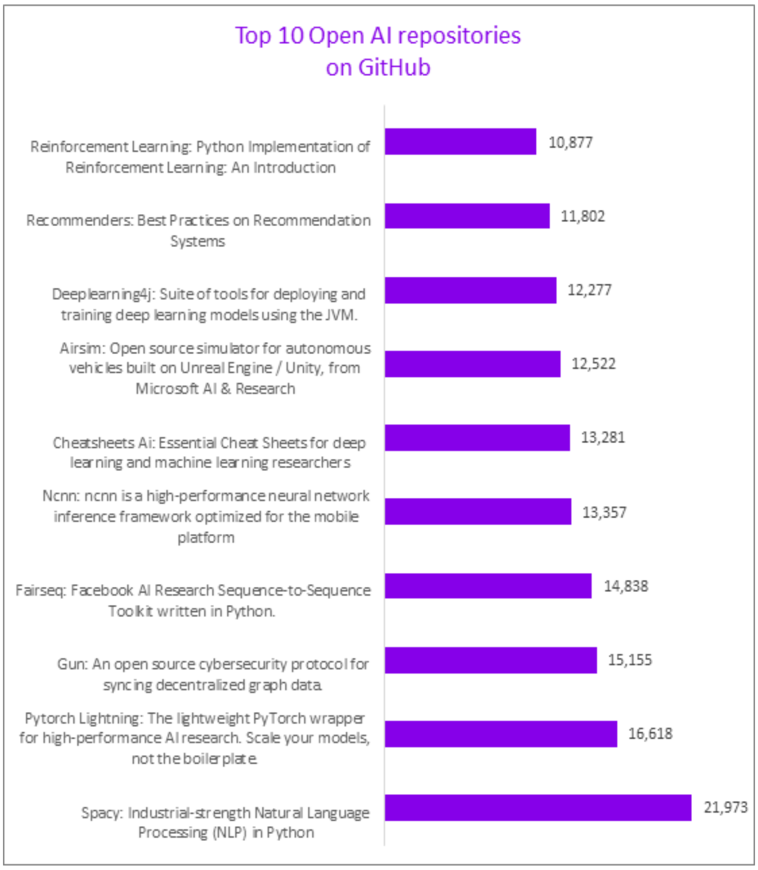


***Figure 109: List of the top 10 AI repositories on GitHub, according to Awesome Open Source*573.**

Openness in information technology (IT) is referred to as a feature of an IT system. Such a system is characterised by interoperability, portability, and extensibility. These can be implemented using interfaces, standards, and IT architectures. Next to these technical aspects, openness is also based on partnership between the involved partners (IT customers, IT vendors and/or IT service providers).[574] Schlagwein et al. define openness as "[...] often deeply embedded in information technology (IT) and [as] both a driver for and a result of innovative IT."[575]

Open source is source code which is freely available, modifiable and redistributable. The distribution terms of open-source software are handled by licenses (e.g., Apache License, 2.0 or MIT License). **Open access (OA)** refers to open and free-of-charge access to scientific publications for all entities worldwide. OA also means that research results are ideally flexibly

---

573 Awesome Open Source: https://awesomeopensource.com

574 Steven, Vettermann. "Code of Openness". CPO. ProSTEP iViP. Retrieved 10 January 2017.

575 Schlagwein, D., Conboy, K., Feller, J. et al. "Openness" with and without Information Technology: a framework and a brief history. J Inf Technol 32, 297–305 (2017). https://doi.org/10.1057/s41265-017-0049-3

reusable. OA publications are always electronic or published online, which may appear in printed form later on.[576]

A **repository** is a storage location for software. It usually includes a table of contents and code documentation. It is typically under version control and serves as a working directory for software development, often related to a specific project or software module.

Democratisation is the action of making something available to everyone. This includes introducing democratic systems and principles. Recently, many major tech companies, e.g. Google or Microsoft, have made the democratisation of AI one of their major goals. At this point it is still unclear how this claim differs from e.g., the open access movement.[577] However, it is clear that the democratisation of AI must champion for broader participation. This concern stems from today's reality: the ability to engineer and make use of AI technologies rests in the hands of a privileged few.[578] To disrupt the workings of today, the barriers to participating in AI development need to be reduced.

Research challenges

Today, AI technology is largely only accessible to advanced users – users that know how to code, how to access repositories, and have the knowledge to apply code to their specific use cases. On the one hand, low tech companies wanting to apply AI might lack the knowledge of what the technology can do for them or lack the AI talent or financial means to employ them. On the other hand, AI savvy companies fear losing their intellectual property and restrict the accessibility to code, models, and data. Bridging the gap and championing democratisation and participation will be one of the major tasks to advance AI in a way that it becomes a common good. More specifically, this task will have to address the following challenges:

Suitable general APIs. First, the AI tools in any modular toolbox must implement suitable APIs. These APIs are often designed with a specific application in mind. On the other hand, the tools should be reasonably versatile, that is, it should be possible to reuse them for other, and future, applications. For this purpose, the implemented APIs must be sufficiently general. What, for example, should be the format of a speech-to-text transcription? The transcription can surely be represented as a string, but for some applications, it might also be interesting to provide confidence scores for the recognised words. This also leads to the question whether formats which are widely used for data exchange between ***microservices*** today such as JSON and XML are sufficient for representing the input and output of AI tools. How can the standardisation process which is needed to synchronise the data formats for a given modular toolbox be organised?

Simplified and configurable modular toolboxes. Second, setting up an actual application by combining AI tools from a modular toolbox is often not just straightforward. Instead, it requires

[576] P. Suber. "Open Access Overview". Retrieved 29 November 2014.
[577] Garvey C. A framework for evaluating barriers to the democratization of artificial intelligence. InThirty-Second AAAI Conference on Artificial Intelligence 2018 Apr 29.
[578] Wolf CT. Democratizing AI? experience and accessibility in the age of artificial intelligence. XRDS: Crossroads, The ACM Magazine for Students. 2020 Jul 9;26(4):12-5.

specific knowledge and skills regarding the used tools, the toolbox or, in other words, the platform for connecting these tools, and the computational environment such as the server or the cluster where the application is to be deployed. These requirements can be a high burden for inexperienced individuals and teams. Therefore, the following questions need to be answered: How can the deployment of AI applications arising from modular toolboxes be simplified? On the other hand, how can very simple "one-click" solutions be designed in a way that they remain configurable for more experienced users, for example with respect to the scalability and the allocated resources of the deployed services?

Legal and ethical concerns. Finally, the usage of modular toolboxes also raises legal and ethical questions. When applications are designed by combining AI tools from modular toolboxes, how can the contributions of all involved developers and researchers be acknowledged? What are the requirements with respect to licensing and data protection? How can the transparency and the security of such an application be ensured? How can its trustworthiness be assessed? Who will be responsible for infringements and damages caused by such an application? Although many approaches have been proposed and tested, these questions have not yet been fully answered.

Societal and media industry drivers

Vignette 1: Streaming pipeline for news live feeds based on open source AI modules

It is the year 2030, Catarina is an editor at a news outlet. Her daily operational task is overseeing the websites' live feeds, e.g., weather report, traffic information, or live tickers for real-time news events. These feeds are automatically created by an AI system. It has been created from an open source repository and it has learned the semantics from millions of weather reports written in English. A streaming pipeline has been established that automatically feeds new input from other news outlets into the model to improve its output. The AI system makes use of social media accounts and, if applicable, includes videos or images from social media users on the live feed, e.g., photos from extreme rain, wind etc. Additionally, all content that is published online has an attached digital ledger that regulates ownership. Therefore, each digital object has an attached license, defining the possibility to (re)use the asset for specified purposes. The AI system might include an open licensed photo of a sunny beach in today's weather feed and is therefore able to clearly identify and notify the owner. If necessary, remuneration is also easily managed through being able to clearly identify how many times a content piece has been used. If content is duplicated without consent by the owner, digital watermarks are included.

Vignette 2: Supporting cinematographic creativity by open AI tools

After two and a half years of studying at the Madrid Film School, Ricardo is about to become a director. For his graduation, he is working on the realisation of a movie which he had in his mind for some years already. The story line has been updated a few times and is perfect now. Ricardo owns a decent camera and a powerful computer for the postproduction of the footage. For reducing the costs, he has borrowed a professional microphone and he has also casted some of his friends. However, the budget is very low, and he cannot afford to pay musicians for composing and recording music for his movie. Furthermore, professional postproduction software nowadays includes AI tools for adjusting the lightning of a scene and even the mimic expressions of the actors and can therefore be very expensive. Luckily Ricardo has learned as part of his curriculum that there is a European open repository for AI tools where he can find what he needs free of charge. He downloads a tool for automatic music composition and an AI tool for special effects. These tools are easy to integrate into his software setup and he has already used them for other projects. As a further benefit, he knows that they come from trustworthy sources and are compliant with the legal requirements such as licensing and privacy protection. Relieved that he does not have to use unlicensed demo versions anymore, Ricardo finishes the postproduction within a few days because the next creative project is already on his mind.

Vignette 3: Using Open Source NLP for news media delivery in any language

Sophia is a developer at a news outlet and is responsible for multiple website features. Users of the website should be able to choose any language to consume any content. For years, the news outlet management was concerned to provide these features due to bias in the AI models. Since last year, Sophia can easily incorporate models from an international cooperation for Open Source NLP to ensure that language is not a barrier for digital content. Sophia interfaces her software to large, already trained AI translation models. Those models were trained on thousands of openly curated language data sources and are easily accessible in the cloud. The cooperation ensures large, diverse, and well curated data sets for models to be trained on. The respective open repositories for code, data and models have governance mechanisms in place that regulate the shared resources to meet the trustworthiness guidelines, e.g. model interpretability, high model performance, minimised bias and transparency to users through open documentation and system architecture design. Hereby, she follows her company's media guidelines for trustworthy content.

Future trends for the media sector

Modular tool boxes are the future of open AI repositories. The core tasks are always the same, for the media sector e.g., keyword extraction, named entity recognition, face recognition, object detection. These tools can then be connected in series and adapted to the intended use. A difficulty lies in the definition of interfaces between the modules. On the one hand, they should be specific enough to be able to use the functionality of the respective tool, and on the other hand, they should be general enough so that the tool can be deployed in a multitude of workflows.

In connection with the modular tool boxes, the trend is moving in the direction of ***open source software***, and even more specifically in the direction of software without copyleft and without restrictions on commercial use. A recent study from the European Commission highlights that procuring open source software instead of proprietary software could increase the digital autonomy of the public sector. This could reduce total cost and avoid vendor lock-in effects.

Moreover, the study predicts that an additional 0.4% - 0.6% of GDP would be generated, if contributions to open source code are increased by 10%.

The interplay of open repositories and open source software will lead to the **democratisation of AI technology** that is accessible for everybody no matter the technological skills or tools.

The European strategy for data foresees the creation of **interoperable data spaces** in strategic sectors[579], including the media sector. The common European data space for media will ensure interoperable and easy access to key datasets. It will provide an ecosystem for the creation of solutions, tools and models for the creation, curation and distribution of media content[580].

Goals for next 10 or 20 years

Following the arguments from the previous subsections, the main goal for the years to come is the establishment of an ***open repository for AI tools*** which can be used by the media, but also by other sectors. The tools within this repository should cover a ***wide range of AI applications*** and they should also continuously make the most recent research available.

One example for a trailblazing open repository is the AI4EU Experiments[581] platform from the AI4EU project[582] (see Figure 110), funded through the European Commission's Horizon 2020 program. As of today, the platform has more than 200 onboarded AI models and provides functionality for matching models. When creating pipelines, the software automatically provides info for matching ingoing and outgoing pipelines, making it easy for users to create AI pipelines from scratch.

The platform is the cornerstone of a future ecosystem for AI tools and will be growing with contributions from all European-funded AI projects. It is also open to all market participants who wish to contribute technologies. Regarding the provided AI tools, the focus should be on open source software in order to ensure transparency, availability, and sustainability. However, commercial tools should not be excluded from the repository.

[579] Common European Data Spaces: http://dataspaces.info/common-european-data-spaces/
[580] European Commission, Staff working document on data spaces (23 Feb. 2022): https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces (pp. 36-37)
[581] AI4EU Experiments Platform: https://aiexp.ai4europe.eu/
[582] AI4EU - The European AI on Demand Platform (funded by H2020 under grant agreement no 825619): https://www.ai4europe.eu

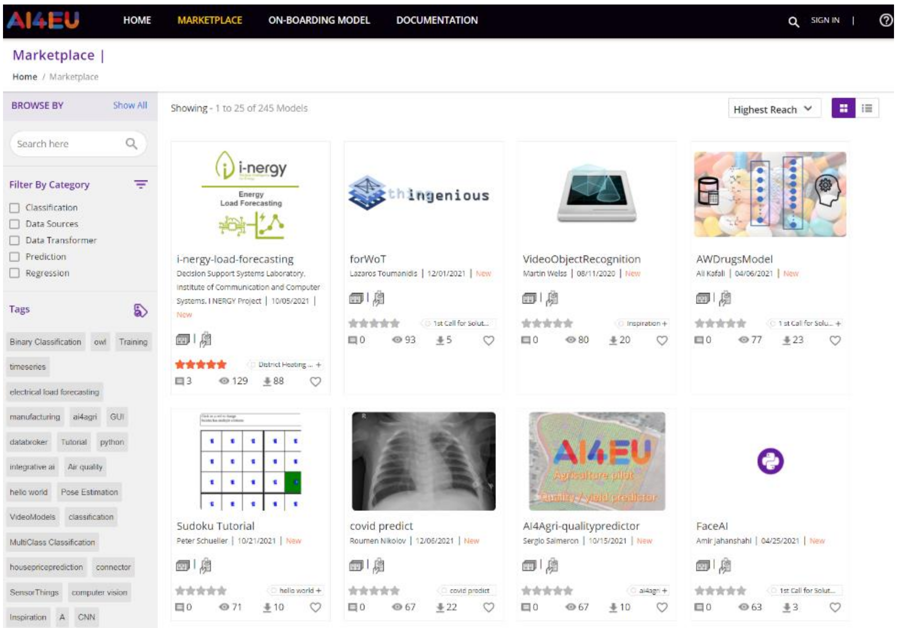


Figure 110: Screenshot of the AI4EU Experiments platform.

11 AI risks & AI policy initiatives: A discussion on ethical challenges, research risks, and regulation frameworks for the media sector

In this section, we dive into the social and ethical implications of AI. We examine ethical, societal, environmental and economic concerns, including bias and discrimination, media (in)dependence, inequality in access, privacy, transparency, accountability, liability, labour displacement, misinformation as an institutional threat, and environmental impact. In addition, we provide a brief overview of existing EU policy and legal initiatives and their impact on future AI research for the media industry.

11.1 Ethical, societal, economic, environmental concerns & risks – a look into the future of AI for the Media

***Contributors**: Anna Schjøtt Hansen (UvA), Tobias Blanke (UvA)*

In AI4Media deliverable D2.2 "*Initial white paper on the social, economic, and political impact of media AI technologies*", an overview of the possibilities and challenges that AI poses for the media sector across the media cycle were presented based on an extensive literature review of both industry reports and scholarly journal articles. In the white paper, the most prevalent societal concerns and risks regarding AI for media are identified, which include:

- Concerns of biases and discrimination;
- Concerns of media (in)dependence and commercialisation;
- Concerns of increased inequality to access to AI;
- Concerns of labour displacement, monitoring and profession transformations;
- Concerns of privacy, transparency, accountability and liability;
- Concern of manipulation and misinformation as an institutional threat;
- Concerns of environmental impacts of AI.

In the following subsections, these concerns are briefly summarised with a clear look towards what practises will be important to counteract potential negative societal impacts of AI for the media industry.

11.1.1 Concerns of biases and discrimination

The risk of ***biases*** and how such biases might lead to ***discriminatory practices*** remain a reoccurring concern across uses of AI in the media sector. AI is here often discussed as a double-edged sword by, on one side often being seen as tools to mitigate both conscious and unconscious biases in human judgement and decision-making, such as mitigating existing media biases. Thereby, offering positive societal impacts relating to, for example, more diversity in coverage, which could improve the public debate and political awareness of previously overlooked societal issues. On the other hand, AI is also built by humans who make decisions on what data to include in the training dataset (which might reflect existing societal biases) and how to design the AI system (e.g., by using standard algorithmic models or deciding on including certain metrics), which can replicate or even enhance existing biases by reinforcing certain ways of 'knowing' and 'seeing' in these systems[583,584].

In the media sector, the implications of biases are perhaps less severe in their immediate effects compared to other sectors (e.g., law enforcement, economic or public sector) as they do not have direct consequences for the individual. However, AI could induce long term negative social and political impacts by, for example, maintaining certain ***gender representations and racially discriminatory patterns*** in the coverage, which could lead certain societal groups to feel underrepresented in the media landscape and disconnect from the public debate as well as produce a skewed portrayal of certain societal topics.

Equally, problematic feedback loops from recommender systems could produce the risk of very individualised and closed off media consumption patterns (i.e. "filter bubbles"), which could in the future negatively affect ***political fragmentation and polarisation*** or support the circulation of radicalising online content. Last, the proved biases and discriminatory effects in content moderation systems, for example on Facebook, as well as the limited availability of such tools in certain languages could prove to have highly negative impacts on the public debate online and contribute to highlighting already dominant voices and further ***undermine minority voices*** online (e.g., racial or gender). Long-term effects that are critical since media accounts, while increasingly contested, remain considered as representations of 'the reality' by many of its audiences[585]. AI induced biases and discrimination pose a serious long-term issue for media institutions that cannot be left for the future. Rather mitigative measures must be put in place. Here, we turn to the general discussion on how to minimise the negative impacts of biases (while of course never completely unavoidable, just as with humans).

The following practices are highlighted as **mitigative measures of biases and discrimination** as a result of AI in the media sector:

- **The need for more domain and social and/or cultural expertise in the development process of AI systems for media.** All AI projects in the media sector should strive for diversity in the team (e.g., in terms of backgrounds, ethnicities or gender) to ensure that the decisions made regarding datasets, classification or metrics are made on a well-founded and reflective basis. Critically, domain knowledge should be prioritised together with social and cultural knowledge in qualifying these decisions.
- **The need for more awareness of the biased 'human nature' of AI amongst media professionals and audiences.** To counteract effects of an overly belief in 'algorithmic objectivity', it will be important to create more critical awareness of the potential of biases in AI systems amongst both audiences who as a baseline find AI more credible and amongst media professionals who also continue to sustain the idea of neutrality in AI.

[583] Campolo, A. et al. (2017) AI Now Report 2017. AI Now Institute at New York University. Available at: https://ainowinstitute.org/AI_Now_2017_Report.pdf

[584] Littman, M.L. et al. (2021) Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report | One Hundred Year Study on Artificial Intelligence (AI100). Available at: https://ai100.stanford.edu/2021-report/gathering-strength-gathering-storms-one-hundred-year-study-artificial-intelligence

[585] Reese, S.D. and Shoemaker, P.J. (2016) 'A Media Sociology for the Networked Public Sphere: The Hierarchy of Influences Model', Mass Communication and Society, 19(4), pp. 389–410. doi:10.1080/15205436.2016.1174268

- **The need for new best practises on how to produce equitable AI use in the media sector.** Currently, the examples of AI projects promoting data justice are scarce. If the sector is to begin a conversation on ways to achieve this, examples of best practises will be needed. This could be in the form of industry research collaborations.
- **The need for sector specific open-source and non-commercial datasets for training AI systems**. As many AI projects today rely on (commercial) open-source datasets, another way to mitigate the potential negative impacts of biases in AI systems is through the development of open-source domain specific datasets, which have been critically examined by a diverse team.
- **The need for best practises and policies of 'diversity by design'**. Currently, limited knowledge and best practices exist on how to evaluate whether, for example, a recommender system is successful – not only in a commercial sense. It would be beneficial to have best practices on how to make such decisions without benchmarking with, for example, purely commercial actors and how to include domain specific measures of diversity in the projects (e.g., filling the gaps of user knowledge etc.). This could also be illustrated through concrete policies on diversity by design.

11.1.2 Concerns of media (in)dependence and commercialisation

Another concern that runs across the different media industry cycles is how the use of AI might induce an ***increased commercialisation of media organisations at the expense of societal responsibility*** as they become more deeply embedded into the platform economy[586,587]. This is not to state that media organisations have not always had a commercial side; private media organisations are a business and PSM's still need to provide legitimisation for their funding by, for example, illustrating their viewership. However, historically these two parts of media organisations have been separated, but over the last 50 years that separation has crumbled[588,589]. AI has proved to further intensify this classic conflict between editorial and commercial side of media organisations by, for example, pushing the limits of the data tracking practises pursued in media organisations[590] or by shifting power to commercial departments who more unquestioned than previously can affect how decisions are made through their knowledge of the infrastructures (e.g., AI or data)[591].

[586] Lindskow, K. (2016) Exploring Digital News Publishing Business Models: A Production Network Approach. PhD Thesis. Copenhagen Business School. Available at: https://research.cbs.dk/en/publications/exploring-digital-news-publishing-business-models-a-production-ne
[587] Sørensen, J.K. and Van den Bulck, H. (2020) 'Public service media online, advertising and the third-party user data business: A trade versus trust dilemma?', Convergence, 26(2), pp. 421–447. doi:10.1177/1354856518790203.
[588] Willig, I. (2010) 'Constructing the audience: a study of segmentation in the Danish press', Northern Lights: Film & Media Studies Yearbook, 8(1), pp. 93–114. doi:10.1386/nl.8.93_1
[589] Willig, I. (2021) 'From audiences to data points: The role of media agencies in the platformization of the news media industry', Media, Culture & Society. doi:10.1177/01634437211029861.
[590] Turow, J. (2016) Media Today: Mass Communication in a Converging World. 6th edn. New York: Routledge. doi:10.4324/9781315681726.
[591] Schjøtt Hansen, A. and Hartley, J.M. (2021) 'Designing What's News: An Ethnography of a Personalization Algorithm and the Data-Driven (Re)Assembling of the News', Digital Journalism, 0(0), pp. 1–19. doi:10.1080/21670811.2021.1988861.

Some of the potential impacts discussed regarding media organisations concerns how the increased valorisation of data and particularly audience data might ***impoverish the overall media landscape by affecting what forms of media content is produced***, by for example valorising certain genres (e.g., more sensationalist content) and de-valorising content that is not 'clickable' but of societal importance – also due to the importance of content circulating well on, for example, social media or ranking high in Google News. Another way to view the potential negative impacts of particularly platformisation is discussed in relation to how it places ***immense power in the hands of very few companies***[592] – in a European context these are the five 'tech giants' Google, Facebook, Microsoft, Apple and Amazon who predominately provide both data and other technical infrastructures for, among other, the media sector. The potential negative impacts of this power imbalance are discussed widely as it places enormous amounts of societal influence in the hands of a few commercial actors, who do not bear a societal responsibility beyond upholding legislation.

Beyond providing concrete infrastructure and providing important intermediary functions in content distribution, these tech giants are also becoming increasingly vital economic patrons in providing support for digital innovation in the media sector. Both Google and Facebook have developed funding schemes in support of innovation, namely the Facebook Journalism Initiative (FJP) (now Meta Journalism Project) and first Google's Digital News Innovation Fund (DNI), followed by Google's News Initiative (GNI). All promising to further the digital innovation at media organisations and ensure a sustainable future for the sector. These innovation programs further strengthen the economic and technological dependence between media organisations and the platforms or digital intermediaries and potentially threaten the ***independence of media organisations***, putting at risk their societal accountability function.

The following practises are highlighted as **mitigative measures of increased dependence and commercialisation** as a result of AI in the media sector:

- **The need for responsible domain specific infrastructures to support the development of responsible AI.** Due to the high reliance on commercialised and platform infrastructures in the development of AI in the media sector, it will be important to develop alternative infrastructures, such as the European Media Data Space[593,594], that perhaps better accommodate the European values and are specific to the media sector.
- **The need for a critical awareness of economic 'patrons' of the media sector and how they affect the development in the media sector.** Currently, limited research exists on the role of 'media patrons' and how they affect the future of the media sector. It will be important that more research is conducted, but also that researchers in fact can get access to these processes, as that is currently highly difficult.

---

[592] Bird, E. et al. (2020) The ethics of artificial intelligence: Issues and initiatives. EPRS | European Parliamentary Research Service. Available at: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/634452/EPRS_STU(2020)634452_EN.pdf

[593] European Media Data Space Project: https://europemedialab.eu/media-policy-europes-media-lab/european-media-data-space-project

[594] European Commission, Staff working document on data spaces (23 Feb. 2022): https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces (pp. 36-37)

- **Funding schemes oriented towards EU values.** In order to counteract the growing role of (US based) platforms in stimulating development, it will be important to develop similar funding schemes that better encompass EU values and the societal function of media.

11.1.3 Concerns of increased inequality in access to AI

Another related concern regarding AI in the media sector is the ***inequalities in access to AI*** solutions by users and AI infrastructure by media organisations. We start with the latter as it more directly relates to the discussion above. Charlie Beckett in a report from 2019[595] highlights how AI is unevenly distributed in the media sector, where particularly local and regional media with smaller budgets are lacking behind, which can reinforce the existing inequalities in the media sector. The inequalities relating to AI could further amplify this trend by further increasing the divide between local and regional as well as niche media organisations and large economically more secure media organisations. As discussed above, AI in fact holds promises to reinvigorate the 'local journalism' through the potential scaling of automated content to cover small events and sports. However, this requires that regional and local actors break the barrier of gaining access to such tools.

Alexander Fanta and Ingo Dachwitz in a report from 2020[596] also show how the funding by the Google DNI Fund, at least in the German context, is oriented towards already large (commercial) media organisations, while smaller start-ups, niche or non-profit organisations are less funded, illustrating how they perhaps selectively stimulate the innovation, making the rich richer, rather than diversifying the access to AI. Again, illustrating how this powerful intermediary elite can stimulate the access to AI in a certain direction, amplifying existing inequalities in the sector[592]. This increasingly uneven access to AI could have serious social and economic impacts in society by ***diminishing diversity in the available media offering*** as certain media organisations unable to leverage the power of AI might be unable to remain competitive.

We now return to the first point mentioned, namely how access to AI services by users is also ***highly unequal across languages***. Training data, for example, often only exist in English or tools are predominately being developed for English. The tools available in other languages also often perform significantly worse due to inherent biases in the datasets. The benefits of AI are, therefore, not shared across the globe, and particularly the divide between the global North and South is growing with the increased use of AI across all sectors and therefore also when it comes to media[497]. The social and political implications of this are vast regarding several AI applications for media, for example, content moderation in large diverse countries like India will suffer under this as they generally will have less efficient content moderation and it is to be expected that many minority languages will experience worse performance, potentially keeping them from partaking in the public debate. There will be a need to place increased focus on this inequality

---

[595] Beckett, C. (2019) New powers, new responsibilities: A global survey of journalism and artificial intelligence. Polis, London School of Economics and Political Science. Available at: https://blogs.lse.ac.uk/polis/2019/11/18/new-powers-new-responsibilities

[596] Fanta, A. and Dachwitz, I. (2020) Google, the media patron. How the digital giant ensnares journalism. preprint. SocArXiv. doi:10.31235/osf.io/3qbp9

of access to ensure that media diversity is sustained and that the benefits of AI become shared, not only by those who are already in privileged positions.

The following practises are highlighted as **mitigative measures against the increased inequality in access to AI** in the media sector:

- **The need for funding schemes and initiatives focusing on media diversity**. It will be important to counteract the trend in private funding schemes where established media organisations remain the main beneficiaries of funding for innovation. To not further the increasing competitive divides in the media sector, funding should be specifically oriented towards furthering media diversity.
- **The need for an increased focus on global AI divides and their consequences**. In general, more knowledge is needed on the severity of the AI divide between the global North and South. It will be important to explore the extent of the issue and its implications further.
- **The development of AI models for diverse languages or language adaptive models**. In order to improve the overall access to AI benefits, AI models for large foreign and minority languages should be developed together with adaptive models that can be more efficiently reused for other languages.

11.1.4 Concerns of labour displacement, monitoring and profession transformations

One of the most discussed impacts of AI has been regarding the labour market and the ***prospect of mass job loss*** when tasks become increasingly automated[583], with sometimes very high estimates of the job losses to be expected, such as 38 million people in the US being in high risk of losing their job because of automation[592]. While the fear of displacement has become more nuanced, both since full automation of many jobs still lies far in the future[583] and as studies have shown that while some jobs will disappear, others will emerge with the growing AI industry[592]. There is also a palpable fear of displacement in the media sector and a few examples of how AI had in fact led to layoffs of 'human' staff. However, in the media sector the fear of displacement discussion has also been nuanced to focus more on the changes AI might impose on the profession and how to maintain the legitimacy of the profession.

One of the impacts of AI has been an emphasised ***focus on 'technical' or 'data oriented' media professionals***[597,598], so an upskilling, rather than displacement, but also how the technology and data focus is increasingly legitimised through managerial shifts[599] and results in media professionals becoming more 'disposable' compared to employees with technical skills who are often much harder to recruit[591]. This is, therefore, producing new asymmetries in the labour market of the media sector, where certain types of jobs and skills are being devaluated. This

[597] Lewis, S.C. and Usher, N. (2013) 'Open source and journalism: toward new frameworks for imagining news innovation', Media, Culture & Society, 35(5), pp. 602–619. doi:10.1177/0163443713485494.
[598] Lewis, S.C. and Usher, N. (2014) 'Code, Collaboration, and the Future of Journalism', Digital Journalism, 2(3), pp. 383–393.
[599] Young, M.L. and Hermida, A. (2015) 'From Mr. and Mrs. Outlier to Central Tendencies', Digital Journalism, 3(3), pp. 381–397. doi:10.1080/21670811.2014.976409.

might require new policies (economic and social) to address this potential societal gap created due to AI[583].

Beyond, impacting what skills are considered important, AI has also been a contributing factor, as part of the overall ***datafication of the media sector***, to the increased importance of data in performance evaluations and recruitment processes, as highlighted by Angele Christin[600]. In the AI Now Institute report from 2017, this is highlighted as a general negative social impact of AI, as the increased reliance on such, often non-transparent, tools have impoverished the working conditions in many places by, for example, ***amplifying the power asymmetry between employer and employee*** or by impacting recruitment processes[583]. These professional changes can negatively affect mental health through, for example, dynamic data visualisations and reminders of goals, which might place stress on the individual to perform or even overperform, due to its importance for keeping one's job[583]. These systems are also generally developed with the employer and not employee in mind, placing the impacts mainly on the individual[601]. Currently, little research is available on the degree of this problem in the media sector, where the focus has more been more on how AI affected production and distribution patterns of media content and not on how AI-enhanced datafied work practices affected media professionals.

The following practises are highlighted as **mitigative measures against the concerns of labour displacement, monitoring and profession transformation** in the media sector:

- **The need for more research and policies addressing potential displacement patterns as a result of AI.** As the increased reliance on AI might result in certain jobs disappearing (e.g., routine tasks) in the media sector as well as across other sectors, providing a societal problem of unemployment. It will, therefore, be important that societal mechanisms and policies are developed to handle the citizens who will be left jobless and in need of specific upskilling.
- **The need for an increased focus on data and AI in media education.** The changes in the media professions also require action from the educational sector who must support students in developing the right skills for the labour market, including increased skills in data science and in understanding how AI systems work.
- **The need for more research on how AI is changing labour conditions and power asymmetries in the media sector.** It will be important to understand how the introduction of AI is enhancing already increasing workplace asymmetries, for example, through the use of performances measurements and with what impacts on the individual and society.

11.1.5 Privacy, transparency, accountability and liability

A plethora of new concerns regarding AI relate to the ***users' rights to both privacy and transparency*** in 'who' they are interacting with, but also to how the introduction produces new

[600] Christin, A. (2018) 'Counting clicks: Quantification and variation in web journalism in the United States and France', American Journal of Sociology, 123(5), pp. 1382–1415.
[601] Crawford, K. et al. (2019) The AI Now Report 2019. AI Now Institute at New York University. Available at: https://ainowinstitute.org/AI_Now_2016_Report.pdf

questions of ***accountability and liability*** was brought forward in the above review. A discussion that is also echoed in the wider discussions of AI[602,603,583,601,592]. The potential social or economic impacts that tracking practices might have on individuals have been highly discussed, for example, in relation to how facial recognition technologies, can allow the identification of individuals across contexts - and even their moods and sexual orientation[592,602], raising questions of individuals' right to privacy in public spaces. Or when the first cases of people being fired based on, for example, GPS data or when data from a pacemaker was used to geographically locate a citizen, leading to his conviction of arson, as a court case recently set precedence for[592]. All these uses of the increasing amounts of trackable personal data have raised serious questions of how to protect the data rights of individuals.

Much like the above discussion on biases, the individual impacts of the data tracking practices related to the use of AI in the media sector, might be much less severe than the examples given above. However, ***untransparent and potentially excessive tracking practices*** by media organisations could have fatal consequences for the trust in these organisations, impeding them from fulfilling the societal task of providing public information. It will, therefore, be important to increase transparency in data use by media organisations, where an overview of the use of data is made easily available and more important the user can react based on this data and retract their consent for certain uses.

Article 52 from the upcoming AI Act points to the right of individuals to know whether they are interacting with an AI system. This is again a wide-reaching discussion that both emphasises the right to know if it is an AI, but also, for example, in the context of healthcare, whether one can request not to have an AI involved in the process of diagnosis[604]. For the media sector, as discussed, the disclosure practices are highly differentiated amongst media organisations, illustrating the need for more harmonisation of how media organisations should approach this new challenge. This includes ***transparency in disclosing when an AI has been involved in the process*** of producing or curating content, but also in how the system came to its decision. This will be important for users, citizens, media and researchers to be able to ***hold the systems accountable*** in cases of harmful decisions.

Another related question that is highly discussed is the question of ***liability regarding AI systems***, because one thing is disclosing that an AI, for example, produced a piece of content, but it is another to determine liability, because many new actors are now involved in this question (e.g., external service providers, in-house developers etc.). Currently, there are still no clear policies or guidelines on this question, which could have negative impacts on media organisations or individual media professionals. Equally, as Lewis et al.[605] point to, there is also a risk that the current regulation will allow a loophole for AI produced content in the case of, for

---

[602] Whittaker, M. et al. (2018) AI now report 2018. AI Now Institute at New York University New York. Available at: https://ainowinstitute.org/AI_Now_2018_Report.pdf

[603] Ada Lovelace Institute (2021) Algorithmic accountability for the public sector: Learning from the first wave of policy implementation. London, UK: The Ada Lovelace Institute. Available at: https://www.adalovelaceinstitute.org/report/algorithmic-accountability-public-sector/

[604] Ploug, T. and Holm, S. (2020) 'The right to refuse diagnostics and treatment planning by artificial intelligence', Medicine, Health Care, and Philosophy, 23(1), pp. 107–114. doi:10.1007/s11019-019-09912-8.

[605] Lewis, S.C., Sanders, A.K. and Carmody, C. (2019) 'Libel by Algorithm? Automated Journalism and the Threat of Legal Liability', Journalism & Mass Communication Quarterly, 96(1), pp. 60–81. doi:10.1177/1077699018755983.

example, personal deformation suits (i.e., lawsuits regarding a false statement made about a person). It will, therefore, become highly important to develop more clarity for media organisations on how to act on this question, and how to translate any legal accountability obligations into organisational practices and internal divisions of responsibility between editors, journalists, data, and economic departments.

The following practises are highlighted as **mitigative measures against the concerns of privacy, transparency, accountability and liability** regarding AI in the media sector:

- **The need for more best practises of responsible data practices in the media sector**. As the extensive use of data continues to grow in the media sector, it will be vital that new best practises are developed to support responsible data strategies that protect the rights of the individual.
- **The need for best practices and policies regarding disclosure of AI systems for the media sector**. As the question of who produced or curated an article is no longer limited to, for example, journalists, editors, and producers, it will be vital that new guidelines for how to disclose the utilisation of AI in these processes are developed to protect the individual's right to transparency.
- **The need for explainable AI solutions that can help users understand how the AI system works and makes its decision**. As users increasingly are partly serviced by AI systems in their media experience, it is important that they have access to understandable explanations of what the system does and on the basis of what data, to uphold their right to, for example, object to the way the decision was made.
- **The need for clearer regulation and guidelines on the liability question regarding AI**. There is a need to help media organisations navigate the liability question that arises from the use of AI systems.

11.1.6 Manipulation and misinformation as an institutional threat

There is also a growing concern amongst the media organisations regarding ***manipulation of content and misinformation***. While this was not related specifically to their own work, the negative impacts of the growing amounts of misinformation were seen as highly detrimental to the ***trust in the media sector***, as for example evident in the survey by Georg Rehm[606]. Making it an institutional threat to the existing media landscape, whose legitimacy is increasingly contested as part of this development. Equally, the fact checking genre and independent fact checking institutions as a result of this growing problem also become a new part of the media landscape. This discussion is, therefore, also slightly different as many of the AI systems that are utilised to mitigate such misinformation are developed by social media platforms or to assist fact checking organisations.

[606] Rehm, G. (2020) The use of artificial intelligence in the audiovisual sector: concomitant expertise for INI report : research for CULT Committee. LU: EU Publications. Available at: https://data.europa.eu/doi/10.2861/294829 (Accessed: 24 November 2021).

The increasing focus on removing misinformation with the assistance of AI systems also raises important discussions regarding ***freedom of expression***, as new guidelines for appropriate forms of censorship must be discussed as well as the potential risks of false positive and negatives in these processes and the lack of complaint mechanisms or satisfactory explanations of why content was deleted[607,608]. Equally, as the practice of fact checking, and particularly AI assisted fact checking grows, these practises must also be more explored, as this remains a highly subjective practice, but which is gaining societal importance. Here both the need for more transparency in the workings of the AI systems used to identify misinformation will become important, particularly as they become intertwined with fact checking organisations through strategic partnerships, such as the ones initiated by Facebook. The topic of disinformation has been discussed in section 9.4 of this deliverable.

The following practises are highlighted as **mitigative measures against the increased threat of manipulation and misinformation** for the media sector:

- **The need for mitigative and adaptive AI systems to counteract misinformation**. To protect the legitimacy of media organisations and the integrity of the online deliberative spaces, it will be important to develop AI systems to assist in content moderation and fact checking efforts. These must be highly adaptive to be effective and counteract adversarial tactics by groups who spread misinformation.
- **The need for more transparency in moderation systems and AI fact-checking systems**. Currently, the AI systems used to identify misinformation on social media platforms remain non-transparent in their workings and the people who experience consequences do not always have access to a satisfying explanation of why, for example, their profile was deleted or to a complaint mechanism. As many fact checkers are today part of strategic partnerships with Facebook, the need to be transparent will become even more important to sustain legitimacy in these institutions that now serve an important societal function.
- **The need for more knowledge on fact-checking as a social practice and its effects in the deliberate landscape**. As fact-checking becomes an important societal function in societies, it will be important to gain more in-depth knowledge in how they construct 'factual' accounts as well as what the consequences of potentially countering epistemologies of the truth might mean for the deliberative space and societal polarisation.

11.1.7 Environmental impacts of AI

A concern that has not really been touched upon but is becoming increasingly important in the more general discussion of AI is the environmental impacts of AI. This was for example highly on the agenda at the 2021 Global Partnership of AI summit in Paris. Training AI systems is 'computation heavy' and leads to large amount of carbon emissions; the storage of data is also an energy intensive process while the hardware needed requires large amounts of natural resources[592]. While AI is also considered to be one of the key solutions to dealing with climate change by providing the basis for 'smarter' solutions across a range of 'carbon heavy' industries

[607] Llansó, E. et al. (2020) Artificial intelligence, content moderation, and freedom of expression. Annenberg Public Policy Center of the University of Pennsylvania.
[608] Gillespie, T. (2020) 'Content moderation, AI, and the question of scale', Big Data & Society, 7(2), p. 2053951720943234. doi:10.1177/2053951720943234

(e.g., agriculture, energy etc.), the impacts on the environment from AI also needs to be further understood so that it is possible to make sensible decisions on when the benefits of AI will surpass the impact on the environment. This remained limitedly present in the current discussion of AI for media but should increasingly be addressed by media organisations who embark on AI projects. The negative environmental impacts also highlight the need for further open-source AI models that can help more than one media organisation. The issue of the environmental impact of AI is further discussed in section 12.3 of this deliverable.

The following practises are highlighted as **mitigative measures against the concerns of environmental impacts of AI** in the media sector:

- **The need for best practises in mitigating the environmental impacts of AI in the media sector**. To help media organisations consider the environment, some guidelines or best practises should be developed to help guide their development processes.
- **The need for more open-source AI models for media**. To minimise the environmental impacts of AI, more collaborative and open-source AI projects should be developed across the media sector.

11.1.8 Summing up

As shown above, there is a need to take action to prevent potential negative societal impacts from affecting society when developing and deciding upon AI applications for the media sector. This requires mitigative actions that span from developing funding schemes that can induce the development of responsible and value sensitive AI to concrete initiatives in the processes of development and implementation, which must be considered in future AI applications and the surrounding discussions for the future of the sector.

11.2 EU policy and regulatory initiatives and their impact on future AI research for the Media

Contributors*: Noémie Krack (KUL), Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL)*

AI4Media deliverable D2.1 "*Overview & analysis of the AI policy initiatives at the EU level*" provided an overview of the EU policy initiatives on AI and the forthcoming Commission's legislative proposals impacting AI. We suggest consulting this document for a wider overview of AI initiatives on different sectors. This section will focus on the initiatives having a clear focus on the media industry.

Two types of initiatives will be presented, namely the policy initiatives (non-binding provisions) and the regulatory initiatives (leading to the adoption of binding legal provisions). They are presented chronologically. We must underline the fact that legislative proposals are only proposals and may be subject to further modifications (for more information on this, see section 12.1 on the future of legislation and regulation).

11.2.1 The Digital Services Act proposal - DSA (REGULATORY)

Released in December 2020, the proposal on a Single Market for Digital Services (Digital Services Act, DSA) aims to ***harmonise rules on the provision of intermediary services*** in the internal market.[609] The text sets up ***transparency and due diligence obligations*** for intermediary services based on an asymmetric approach. Intermediary services, hosting services, online platforms and very large platforms are the providers tackled by the provisions. Intermediary services cover mere conduit[610], caching[611] and hosting services[612]. Such services include wireless local area networks, domain name system (DNS) services, top–level domain name registries, certificate authorities that issue digital certificates, or content delivery networks that enable or improve the functions of other providers of intermediary services, voice over IP services, web-based messaging services and e-mail services. An online platform is a provider of a hosting service which, at the request of a recipient of the service, stores and disseminates to the public information such as social networks, online market places, app stores, online travel and accommodation websites, content-sharing websites. Very large platforms are platforms reaching a number of average monthly active users in the European Union of at least 45 million (10% of the EU population). This would include for instance big social media providers such as Facebook.

As illustrated in Figure 111, if the provider falls in one or several subcategories (hosting services, online platforms and very large platforms), it will have one, two or three extra layers of obligations applicable to his services.

[609] European Commission, Proposal for a Regulation on the European Parliament and of the Council on a Single Market for Digital Services (DSA Act) and amending Directive 2000/31/EC, 15 December 2020, COM(2020) 825 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2020%3A825%3AFIN

[610] Mere conduit services are services consisting of the transmission of information in, or the provision of access to, a communication network, and include the services of telecommunications operators and internet access providers.

[611] Caching services are services that consist of the transmission in a communication network of information provided by a recipient of the service, involving the automatic, intermediate and temporary storage of that information, for the sole purpose of making more efficient the information's onward transmission to other recipients upon their request.

[612] Hosting services are services consisting of the storage of information provided by, and at the request of, a recipient of the service.

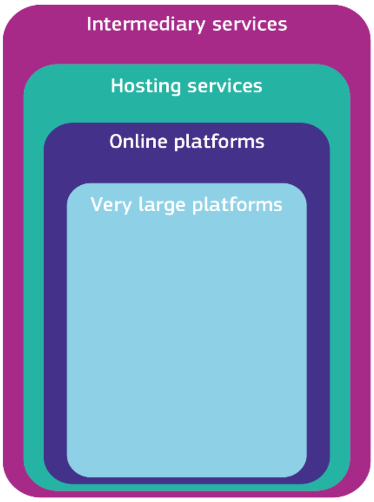


Figure 111: Categories of intermediary services according to the AI Act. Each category corresponds to extra layers of obligations.

The DSA proposal considers the impact of the use of AI based tools used in online media. The preamble of the proposal underlines how algorithmic systems shape information flows online (e.g. via content prioritisation, advertisement display and targeting or content moderation). They also create or may reinforce existing discrimination in content moderation. Below, we discuss how the DSA proposal regulates the use of AI tools in media by different online service providers.

AI and Content moderation

Providers of intermediary services must include in their **terms and conditions** (article 12 DSA proposal), in a clear and ambiguous language, information on any policies, procedures, measures and tools used for the purpose of content moderation, including algorithmic decision-making and human review.

Providers of hosting services shall put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items that they consider to be illegal content. When confirming receipt of the **notice**, they must provide information on the use of automated means for the processing or decision-making of the notice submitted (see article 14 DSA proposal). When communicating their decision to remove or disable access to specific items of information provided by the uploader, hosting providers must include information on the use of automated means in taking the decision. For instance, if the content got detected or identified by the use of automated means (see article 15 DSA proposal).

Online platforms must now put in place an **internal complaint-handling system** for managing the complaints against a decision taken against an information provided/uploaded by a recipient of their services. The decision on the complaint must not be solely taken on the basis of automated means (article 17§5 DSA proposal). The online platforms also get additional obligation when it comes to **transparency reporting** as they must include in their yearly report on content moderation, any use made of automatic means for the purpose of content moderation, including a specification of the precise purposes, indicators of the accuracy of the automated means in fulfilling those purposes and any safeguards applied (see article 23 DSA proposal).

AI and advertising

Online platforms displaying **advertising** on their online interfaces must also ensure that information on the ad is provided in a clear and unambiguous manner and in real time: such as on whose behalf the ad is displayed and meaningful information on the main parameters used to determine the recipient to whom the advertisement is displayed, in other words what are the criteria used for targeting (article 24 DSA proposal).

VLOPS (Very Large Online Platforms) must additionally hold a **publicly available online advertisement repository**. This repository must indicate whether the advertisement was intended to be displayed specifically to one or more particular groups of recipients of the service and if so, the main parameters used for that purpose (see article 30 DSA proposal).

AI and risks assessment

VLOPS must conduct at least once a year a **risks assessment** of the functioning and use of their service which must take into account how their content moderation systems, recommender systems and systems for selecting and displaying advertisement influence any of the systemic risks (see article 26 DSA proposal). Once risks are identified, mitigation measures must be put in place and can lead to adapting content moderation or recommender systems, their decision-making processes, and the like.

AI and recommender systems

The VLOPS shall, in addition to the intermediary providers, set out in their **terms and conditions**, in a clear, accessible and easily comprehensible manner, the main parameters used in their **recommender systems**, as well as any options for the recipients of the service to modify or influence those main parameters that they may have made available, including at least one option which is not based on profiling, within the meaning of the GDPR. Not only providing transparency, VLOPS must also empower recipients to parameter their recommender systems.

AI access for research and enforcement purpose

VLOPS are subject to more stringent enforcement measures when it comes to transparency of automated decision-making systems. They shall **provide access** (see article 31 DSA proposal), on the one hand, to the data that are necessary to monitor and assess compliance with this Regulation to the European Commission or the Digital Services Coordinator. On the other hand, to the data for vetted researchers focusing on identification and understanding of systemic risks created by VLOPS, including their content moderation systems, recommender systems and their systems for selecting and displaying advertisements. This data access regime will be further clarified in delegated acts. However, the provision mentions that only the system for researchers should be in compliance with the GDPR and take into account the interest of VLOPS such as the protection of confidential information, in particular trade secrets, and maintaining the security of their service. Questions remain about the implementation of this regime and about the limitation attached to the restrictive vetted researcher scheme.

The power to conduct **on-site inspections** is being granted to the Commission and it can request VLOPS to provide *explanations on the algorithms* used (see article 54§3 DSA proposal). The Commission can take actions to monitor the effective implementation and compliance with the DSA by **ordering access** and explanations to the VLOPS **databases and algorithms.**

11.2.2 Proposal for a Regulation laying down harmonised rules on artificial intelligence - AI ACT (REGULATORY)

Released by the EC in April 2021, the AI Act proposal aims to ensure the proper functioning of the internal market by setting harmonised rules in particular on the development, placing on the Union market, and use of products and services making use of AI technologies or provided as stand-alone AI systems.[613] A ***risk-based approach*** is pursued by the proposal, which means that specific obligations and requirements are foreseen for distinct categories of AI systems. It includes AI systems creating ***Unacceptable risk*** (they are hence prohibited), ***High-risk*** (requirement specific ex ante, ex post requirements and a supervision and enforcement structure and regime), ***limited risks*** (only requiring transparency obligations) and ***minimal risks*** (free use) (see Figure 112).

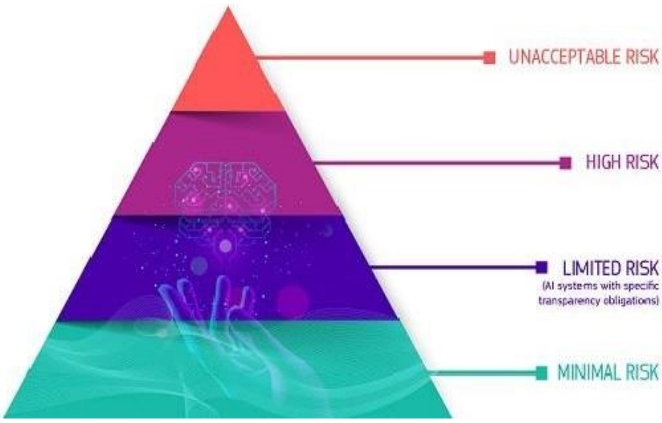


Figure 112: The risk-based approach proposed by the AI Act: AI systems are classified in four categories based on the risk they create.

[613] European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (AI Act) and amending certain Union legislative acts, 21 April 2021, COM(2021) 206 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206

AI media applications are not specifically tackled by the regulation, except chatbots and deepfakes. This has raised concerns from the media community since from the text description, it is not clear whether AI media applications could fall or not in a certain category.

Prohibited Practices. Article 5 of the AI Act proposal prohibits certain practices, such as the use of subliminal techniques or the exploitation of a specific vulnerability of a specific group of persons. Could recommender systems and targeted advertising fall within the subliminal techniques being considered as manipulative systems?[614] It is unclear whether that would be the case, therefore additional guidelines or clarifications would be welcomed.

High risk. 'Media' is not considered as a 'high-risk' area and is not listed in the high-risk list in Annex III. However, research is being conducted on the risks associated to the use of AI in media sector: discrimination, bias, threats to fundamental freedom such as freedom of expression when AI systems are used for content organisation, moderation. One can wonder whether recommender systems could fall under this category. It is likely not for e-commerce, online dating, entertainment content or search engines.

Limited risks. The current wording of article 52 establishing transparency obligations lacks clarity, which could lead to legal uncertainty. Article 52§1 deals with AI systems intended to interact with natural persons. They shall be designed and developed in such a way that natural persons are informed that they are interacting with an AI system. Could this apply not only to chatbots but to recommender systems, automated journalism, as well? Article 52§3 on deepfakes provides that users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deepfake') shall disclose that the content has been artificially generated or manipulated. However, this obligation comes with a few exceptions. Firstly, the provision only applies to users. According to Article 3§4 of the proposal, users are defined as follows: "*Any natural or legal person, public authority, agency or other body using an AI system in the course of a personal non-professional activity shall not be considered as a user*". Secondly, the provision shall not apply where "*necessary for the exercise of the right to freedom of expression and the right to freedom of arts and sciences and for law enforcement purposes*" (article 52§3,2). These limitations are considerably restricting the added value of the transparency requirements for deepfakes as many of the malicious ones out there could be considered covered by one or several exemptions. For researchers, this labelling obligation could be optional when the absence of labelling would be necessary for the exercise of freedom of sciences. It remains to be seen in which situation the obligation could be applicable. In any case, labelling is always a good practice for ensuring trustworthy use of AI in media applications.

[614] M. MacCarty, K. Propp, Machines learn that Brussels writes the rules: the new EU's new AI Regulation, Lawfare Blog, April 2021, https://www.lawfareblog.com/machines-learn-brussels-writes-rules-eus-new-ai-regulation .

11.2.3 Resolution on AI in education, culture and the audiovisual sector (POLICY)

On 19th May 2021, the European Parliament (hereafter EP) adopted a resolution on AI in education, culture and the audio-visual sector.[615] The EP acknowledged the considerable influence of education, cultural programmes and audio-visual content in shaping people's beliefs and values. This influence requires that any development, deployment and use of AI and related technologies in these sensitive sectors must fully respect the fundamental rights, freedoms and values enshrined in the EU treaties.

AI media applications have already entered the entire creation value chain and are present in users' daily lives. This is well reflected by the EP resolution, which makes a non-exhaustive list of AI media applications, their opportunities and challenges (for more information, see AI4Media deliverable D2.1 "*Overview & Analysis of the AI Policy Initiatives in EU level*").

Additionally, the resolution asked the EC to present a ***general regulatory framework***, which applies to all applications of AI, and to complement it with sector-specific rules, for example for audio-visual media services.

MEPs invited the EC to introduce ***strict limitations on targeted advertising*** based on the collection of personal data, starting with a ***ban on cross-platform behavioural advertising***. Incorporating the ***gender and the diversity dimension in datasets, training, education, composition of developers' team, research*** was also requested. Having a framework in place to help determine which content is protected under the IPR legislation would be necessary to solve ambiguities. They called for the establishment of a clear ethical framework for the use of AI technologies in media to prevent all forms of discrimination and ensure access to culturally and linguistically diverse content at Union level, based on accountable, transparent, and inclusive algorithms, while respecting individuals' choices and preferences. Transparency would also be a key tool for all the stakeholders: the media users, the media organisation providing content, the researchers, and the intermediaries. The following questions came into prominence: "*How is the content ranked?; What are the parameters used?; How can the users be empowered to influence to set their preference?*" They pointed that this is only partially achieved in the DSA and that more could be done in this regard (see below).

Focusing on the audio-visual sector, further recommendations were made. Algorithms used by media service providers, video sharing platforms (VSPs) and music streaming services should ensure that ***personalised suggestions do not put forward the most popular works, for targeted advertising, commercial purposes or to maximise profit***. Indicators on content promoting diversity, European work should be included. The EP called both for a recommendation on algorithms and personalised marketing striving for ***explainability, transparency and non-discriminatory outputs*** and on ***user control and empowerment over algorithms used for content recommendation*** with an option to opt-out from recommendation and personalised services. They also put forward that algorithms should only be used as a flagging mechanism in

[615] European Parliament, Resolution on artificial intelligence in education, culture and the audio-visual sector, 19 May 2021, 2020/2017(INI), https://oeil.secure.europarl.europa.eu/oeil/popups/summary.do?id=1663438&t=d&l=en

content moderation, subject to human intervention. Furthermore, access to data held by intermediary providers was underlined.

On AI media applications, disinformation and deepfakes, the EP asked for more visibility, education on the matter. MEPs also requested an ***appropriate legal framework to govern deepfake*** creation, production, or distribution, including automatic tagging, stating that the content was manipulated and strict limitation when used for electoral purposes. Interestingly, EP's requests strive to be way more ambitious than the provision contained in the AI Act (see section 11.2.2). They also urged for more user control over the content they are suggested.

11.2.4 Code of Practice on disinformation (SELF-REGULATION) & EC Guidance to strengthen the Code (EU POLICY)

On **26th May 2021**, the European Commission released its **guidance to strengthen the Code of Practice on Disinformation**.[616] For information, at the EU level, disinformation is currently left to a self-regulation instrument namely **the Code of Practice on disinformation**[617] (hereinafter: the Code). The Code has been signed by major online platforms and members of the advertising industry. **Regarding AI media applications**, the Code indicates that the signatories acknowledge the importance to cooperate by providing relevant data on the functioning of their services, including data and general information on algorithms. They also indicate committing to put in place clear policies regarding identity and the misuse of automated bots. They also commit to indicate in their policies what constitutes impermissible use of automated systems and to make this policy publicly available on the platform and accessible to EU users.

In September 2020, after a year of existence, the Code received an unsatisfactory assessment[618] by the EC. Following these events, the EU increased its regulatory efforts to fight disinformation with several EU initiatives. This includes the DSA proposal, which suggests the introduction of a ***subtle co-regulatory mechanisms for regulating disinformation***. This also includes the EU Democracy Action Plan[619], which announced the publication of the European Commission

[616] European Commission, Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Guidance on strengthening the Code of Practice on Disinformation, 26 May 2021, COM(2021) 262 final. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0262
[617] Code of practice on disinformation, April 2018, https://digital-strategy.ec.europa.eu/en/policies/code-practice-disinformation
[618] European Commission, Staff Working Document – Assessment of the Code of Practice on disinformation – achievements and areas for further improvement, 10 September 2020, SWD(2020) 180 final, https://digital-strategy.ec.europa.eu/en/library/assessment-code-practice-disinformation-achievements-and-areas-further-improvement
[619] European Commission, Communication to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the European Democracy Action Plan, 3 December 2020, COM(2020) 790 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:790:FIN

Guidance to strengthen the Code of Practice on disinformation, inviting signatories of the Code to submit a revised version of the Code.

The Guidance aims to reduce the financial incentives for disinformation actors, empower users and encourage flagging harmful content. It pushes providers of AI-enabled online systems for a ***safe design commitment***, especially in light of risks for users' behaviours using their services. They must also take into consideration the relevant provision of the AI Act (p.12-13). The guidance also specifically calls for ***recommender system accountability*** through transparency measures about the criteria used for content ranking and prioritisation and also the possibility for users to customise the ranking algorithms.

The signatories are now expected to deliver the strengthened Code of Practice by the end of **March 2022**.[620]

11.2.5 Proposal on transparency and targeting of political advertising (REGULATORY)

In November 2021, the European Commission released a new proposal for a regulation focusing on political advertising, transparency and targeting.[621] It will establish harmonised transparency measures to bring uniformity to the diverse regulations adopted in the Member States, enhance trust in the political debate and the integrity of the political debate. The text will bring legal certainty to a booming sector and will prevent the replication the Cambridge Analytica events and the interference and influence on democratic events. Rules for political targeting based on personal data and sensitive personal data will be established. The text will apply to both offline and online political advertising and targeting.

The proposal complements the regime set by the DSA proposal for commercial advertising and applies in compliance with the rules sets in the General Data Protection Regulation (GDPR).[622] The harmonious co-existence of the two instruments is a core aspect of the proposal. The text sets a uniform set of rules for growing practices which come with considerable risks for democracy and fundamental rights protection. The European Commission sets a broad scope for political advertising. The material scope is quite wide. So is the political advertising definition, which includes any message for or on behalf of a political actor which is liable to influence the outcome of an election or referendum, a legislative or regulatory process or voting behaviour

[620] European Commission, Revision of the Code of Practice: the strengthened Code expected by March 2022, December 2021, https://digital-strategy.ec.europa.eu/en/news/revision-code-practice-strengthened-code-expected-march-2022

[621] European Commission, Proposal for a Regulation of the European Parliament and of the Council on the transparency and targeting of political advertising, 25 November 2021, COM(2021) 731 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0731

[622] European Commission, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), 27 April 2016, *OJ L 119,* https://eur-lex.europa.eu/eli/reg/2016/679/oj.

(Article 2§2). The personal scope targets a wide range of stakeholders involved in political advertising: political parties, political advertising services, publishers, sponsors.

Firstly, it sets up ***transparency obligations for political advertising*** services. In short, political advertisements shall be clearly labelled as such and include strict transparency information (who paid for it, how much, for which democratic event, etc.). Periodic reporting on political advertising services would become mandatory. Notification mechanisms for illegal political advertisement should be put in place by advertising publishers. Cooperation is also foreseen between the providers of political advertising services and competent authorities and interested parties such as vetted researchers, civil society organisations, electoral observers and so forth.

Secondly, the proposal introduces ***specific requirements for targeting and amplifications techniques related to political advertising***. Political targeting and amplification techniques would now need to be explained publicly in unprecedented detail (transparency notice, user friendly format and information content, record keeping on techniques and criteria used). The proposal establishes a **ban on political advertising** based on sensitive personal data (article 12§1). Sensitive personal data are data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, or trade union membership, and the processing of genetic data, biometric data for the purpose of uniquely identifying a natural person, data concerning health or data concerning a natural person's sex life or sexual orientation (article 9§1 GDPR). This prohibition is however nuanced by some **exceptions** (article 12§2) including data subject's explicit consent. Some already point out that given the dominant position of giant technological companies, users risk to have little choice regarding their data if they want to use their services, which goes against the concept of free and informed consent.[623]

Additionally, specific schemes should be put in place to ensure that data subjects can effectively exercise their GDPR rights. Recital 51 clarifies that "*The tools made available to the individuals to support the exercise of their rights should be effective to prevent an individual from being targeted with political advertisements, as well as to prevent targeting on the basis of specific criteria and by one or several specific controllers*". The AI systems employed need therefore to be explained in a transparency manner.

Data access provisions are also in place for competent authorities exercising their supervisory tasks (article 15§3). The proposal will now be discussed within the European Parliament and the Council of the European Union. The goal of the European Commission is to have the regulation ready before the EU elections of 2024.

[623] Euractiv, How new, binding EU transparency standards for political advertising could be even higher, 30 November 2021, https://www.euractiv.com/section/digital/opinion/how-new-binding-eu-transparency-standards-for-political-advertising-could-be-even-higher/

11.2.6 OSCE Policy Recommendations on artificial intelligence and freedom of expression (POLICY)

The Organization for Security and Co-operation in Europe (OSCE) has released on 20 January 2022 a manual containing policy recommendations on the most effective ways to safeguard freedom of expression and media pluralism, when deploying advanced machine-learning technologies such as AI.[624] The publication is part of the project "Spotlight on Artificial Intelligence and Freedom of Expression" (#SAIFE).

14 recommendations are addressed to the OSCE Participating States. The manual contains the findings of expert workshops organised in 2021 by the OSCE and Access Now, which unpacked and analysed the main challenges that AI tools pose to human rights, in particular, the right to freedom of expression and opinion, and media freedom and pluralism. It is organised around AI in content moderation and curation. The analysis focuses on challenges relating to the use of AI in these domains such as security threats, hate speech, media pluralism, and surveillance-based advertising.

11.2.7 Conclusion

In conclusion of these policy and regulatory initiatives, we see that AI media applications are on the verge of being specifically regulated in legal instruments. The common approach chosen to deal with them is principally transparency requirements and then, at a level, users' empowerment, especially by providing information on the use of automated means for the processing or decision-making in content moderation.

[624] Organization for Security and Co-operation in Europe (OSCE), The Representative on Freedom of Media, Spotlight on artificial intelligence and freedom of expression, a Policy Manual, January 2022, https://www.osce.org/files/f/documents/8/f/510332_0.pdf

12 What are the forces shaping the future?

In this section, we discuss the forces that could shape the future of the use of AI in the media industry. We focus on legislation/regulation, the pandemic and its impact, and finally on the climate crisis.

12.1 Future legislation and regulation

***Contributors**: Noémie Krack (KUL), Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL)*

The EU legislative agenda on technology-oriented proposals is busy and, as outlined in section 11.2 of this report, it will impact AI media applications and frame their future use. The EU legislative proposals are at different stages of the EU democratic process. AI media applications will become more and more regulated following the trend we saw in the past years. However, the regulation efforts are only in their beginning, and therefore, more will probably come into the picture in the years to come.

12.1.1 The DSA negotiations

The **DSA proposal** is the most advanced technology-oriented proposal of our analysis in section 11.2.

Council of the European Union position

On 25th November 2021, the Council of the European Union officially adopted its general approach (meaning its official position) on the DSA proposal.[625] While expanding the scope of the proposal and clarifying certain aspects in the recitals, the Council did not drastically change the provisions related to AI media applications set in the EC proposal. In brief, their main amendments in relation to AI media systems are the following:

- The text now explicitly includes **online search engines**. They became a new category of intermediary services. Very Large Online Search Engines with more than 45 million of active users will become subject to the provisions applicable to VLOPS.
- Recital 50 a) provides that "*common and legitimate **advertising practices** that are in compliance with Union law should not in themselves be regarded as constituting dark patterns*".
- New article 24 b provides that "*Providers of **online marketplaces** shall not design, structure, or organise their online interface in a way that either purposefully or in effect deceives or manipulates recipients of the service, by subverting or impairing their autonomy, decision-making or choice*".

[625] Council of the European Union, Proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC - General approach, 18 November 2021, 13203/21, https://data.consilium.europa.eu/doc/document/ST-13203-2021-INIT/en/pdf

- Recital 60 specifies that for auditing purposes, the auditor should ***get access to data related to algorithmic systems of VLOPs***. Rec 64 then further adds that the access request for monitoring compliance purposes can also include where appropriate training data and algorithms. Algorithms, which can be further investigated on during on-site inspections (article. 54§1a, d).
- The ***transparency reporting*** for hosting services now includes the number of notices processed exclusively by automated means.
- Recital 54 clarifies that the DSA does not require providers of online platforms or of online search engines to perform specific ***tracking*** of individuals online, nor to discount automated users such as ***bots***.

European Parliament position

The Internal Market and Consumer Protection (IMCO) Committee was the responsible committee for the Digital Services Act proposal but other Committees (ECON, FEMM, ITRE, TRAN, CULT, JURI, LIBE) could provide their opinion between September and October 2021. The IMCO Committee adopted on 13 December 2021 its position on the DSA proposal by 36 votes in favor, 7 against and 2 abstentions. The EP debated and voted on the IMCO's report during the January 2022 plenary session.[626] The European Parliament finally adopted, on **20 January 2022**, its position on the DSA proposal. Last-minute amendments in the plenary vote introduced significant changes to the EC proposal. A selection of amendments will be presented below and their impact on the use of AI in media applications.

AI and Advertising

The EP provided more attention to advertising, the data and the transparency of advertisement than the initial proposal. Targeting or amplification techniques involving the data of minors for the purpose of displaying ads will be prohibited, as well as targeting individuals on the basis of sensitive data which allow for targeting vulnerable groups. Recital 52 insists that recipients of the service can refuse or withdraw their consent for targeted advertising purposes, in a way that is not more difficult nor time-consuming than to give their consent. Furthermore, the amendment specifies that online platforms should also not use personal data for commercial purposes related to direct marketing, profiling and behaviourally targeted advertising of minors. The amendment requires also that more detailed information on advertisement be mentioned in the advertisement repositories.

AI and algorithmic recommender systems

[626] The final amendments adopted by the EP are in the European Parliament Amendments adopted by the European Parliament on 20 January 2022 on the proposal for a regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, https://www.europarl.europa.eu/doceo/document/TA-9-2022-0014_EN.html

The amendments set that not only **VLOPS** (see recital 67) but also **online platforms** (see new recital 52, a) and new article 24, a)) should ensure that recipients can understand how recommender systems impact the way information is displayed and can influence how information is presented to them. **VLOPS** should ensure that at least one option of the recommender system is not based on profiling and that recipients can influence the information presented to them through making active choices (see art. 29§1 and recital 62). In addition, **VLOPS** should implement appropriate technical and organisational measures for ensuring that recommender systems are designed in a consumer-friendly manner and do not influence end users' behaviour through dark patterns.

AI and content moderation

The definition has been clarified and now article 2§1, p) provides that content moderation includes automated or not automated moderation of online content. In terms of liability of intermediaries, article 7 §1 a) now explicitly states that **providers of intermediary services** shall not be obliged to use automated tools for content moderation or for monitoring the behaviour of natural persons. Recital 25 now provides that **they** should make best efforts to ensure that where automated tools are used for content moderation, the technology is sufficiently reliable to limit to the maximum extent possible the rate of errors where information is wrongly considered as illegal content. In addition, **their** terms and conditions must include information on algorithmic decision-making and human review including a qualitative description of whether and how automated tools for content moderation are used in each official language (article 12§1 a). An interesting provision aims to limit the spread of illegal content. For frequent uploaders of illegal content or frequent unsubstantiated complainants, the suspension of the services by **online platforms** shall be ordered taking into account if an automated content recognition system was used to file the repetitive complaints (article 20 §3 da). Suspension can be declared permanent, if the items removed were components of high-volume campaigns to deceive users or manipulate platform content moderation efforts (article 20 §3a).

AI and systemic risks assessment

Article 26§1 specifically indicates that the systemic risks assessment by **VLOPS** is not only stemming from the functioning and use made of their services but also stemming from the design, algorithmic systems, intrinsic characteristics of the services. Therefore the advertising model and their algorithmic systems shall be taken into account for the assessment and for the appropriate mitigation measures (article 26§2 and article 27 §1 a) of the EP amendments. Recital 52 also includes behavioural advertisements in the systemic risks to assess by the online platforms.

AI and access request

VLOPs shall also explain the design, logic and functioning of algorithms if requested to do so for oversight and enforcement purposes (article 31§1 a, recital 64). They also broadened the scope of access for research purposes to non-for-profit organisations or associations (article 31).

AI and deepfakes

A new article (30 a) is introduced and provides that where **VLOPS** become aware that a certain piece of content on their services is a deepfake (video, audio or other file), they shall label the content in a way that informs that the content is inauthentic and that is clearly visible for the recipient of the services.

AI and free decision making

Online platforms should be prohibited from using deceiving or nudging techniques to influence users' behaviour through dark patterns. Indeed, the importance for services recipients to make free, autonomous and informed decisions was underlined and new §1 a) of article 24 and recital 39 a) stated that *"providers of intermediary services should be prohibited from deceiving or nudging recipients of the service and from distorting or impairing the autonomy, decision-making, or choice of the recipients of the service via the structure, design or functionalities of an online interface or a part thereof ('dark patterns')".* This includes the interface design, the visual prominence, the consent option, and so forth.

Following steps

The amendments presented above showed how the European Parliament levelled up the ambition of the text to handle a selection of AI media applications. It added the provisions on deepfakes, on algorithmic accountability and transparency as well as elements in relation to recommender systems and advertising. On the other hand, it is said that for the French Presidency, which presides over the EU Council in the first half of 2022, the governance framework for VLOPs and the inclusion of very large search engines in the scope of the DSA proposal, are two critical issues in the negotiations.[627] The respective positions of the two institutions will be discussed in trilogues reuniting the EC, the EP and the Council in order to find a common voice among the different positions. There are currently five political trilogues scheduled starting already on 31st January 2022.[628]

12.1.2 The AI Act negotiations

The AI Act proposal is still being discussed in the European Parliament and the Council[629]. In November 2021, the Slovenian presidency presented a progress report (draft compromise) on

---

627 Euractiv, DSA: MEPs gear up for negotiations ahead of kick-off trilogue, 26 January 2022: https://www.euractiv.com/section/digital/news/dsa-meps-gear-up-for-negotiations-ahead-of-kick-off-trilogue/

628 For more information on the progress: https://oeil.secure.europarl.europa.eu/oeil/popups/ficheprocedure.do?lang=en&reference=2020/0361(COD), https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-digital-services-act, https://eur-lex.europa.eu/legal-content/en/HIS/?uri=COM:2020:825:FIN

629 For more information on the procedure progress: https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX:52021PC0206 and https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence

discussions held so far within the Council on the AI draft proposal. In Parliament, the discussions will be led by the Committee on Internal Market and Consumer Protection (IMCO) and the Committee on Civil Liberties, Justice and Home Affairs (LIBE) under a joint committee procedure. Since its publication (April 2021), considerable time was taken to allocate the responsibilities within the European Parliament, therefore discussion are about to start within the Committees.

12.1.3 Regulation proposal on transparency and targeting of political advertising

This proposal has recently been released. Therefore, it will first go through the European Parliament and the Council of the European Union in the following months. Both will need to adopt their positions before entering in discussion.[630]

12.1.4 Council of Europe Framework

In terms of regulation, we should not forget the international legal framework. The Council of Europe (CoE) is an active actor on AI related initiatives under the angle of fundamental rights. Recommendations on the human rights impacts of algorithmic systems[631], a declaration on the manipulative capabilities of algorithmic processes[632], and guidelines on facial recognition[633] were adopted in the past few years. The CoE hosted a conference[634] on artificial intelligence and the challenges and opportunities for media and democracy, where they published a background paper about the impacts of AI-powered technologies on freedom of expression[635], and adopted a Declaration and Resolutions on AI tools used for the creation, moderation and distribution of online content[636]. An ad-hoc Committee on Artificial Intelligence (CAHAI) was created in relation to the Council of Europe's effort on human rights. In December 2021, the CAHAI adopted the "Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy and the rule of law", which closes the mandate

---

630 For more information on the procedure progress: https://eur-lex.europa.eu/legal-content/EN/HIS/?uri=CELEX:52021PC0731

631 Council of Europe, Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems, 8 April 2020: https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

632 Council of Europe, Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes, 13 February 2019: https://search.coe.int/cm/pages/result_details.aspx?objectid=090000168092dd4b

633 Council of Europe, Consultative Committee of the Convention for the protection of individuals with regard to automatic processing of personal data, Convention 108, Guidelines on Facial Recognition;, 28 January 2021: https://rm.coe.int/guidelines-on-facial-recognition/1680a134f3

634 Council of Europe, Conference of Ministers responsible for Media and Information Society, Artificial intelligence – Intelligent politics, Challenges and opportunities for media and democracy, 10-11 June 2021: https://www.coe.int/en/web/freedom-expression/media2021nicosia

635 N. Helberger, S. Eskens, M. van Drunen, M. Bastian, J. Moeller for the Council of Europe, Background Paper, Artificial Intelligence – Intelligent Politics Challenges and opportunities for media and democracy, February 2020: https://rm.coe.int/cyprus-2020-ai-and-freedom-of-expression/168097fa82

636 Council of Europe, Final Declaration, Resolution on freedom of expression and digital technologies, Resolution on the safety of journalists, Resolution on the changing media and information environment, Resolution on the impacts of the COVID-19 pandemic on freedom of expression, 11 June 2021: https://rm.coe.int/final-declaration-and-resolutions/1680a2c9ce

of the group for the Council of Europe.[637] The document contains an outline of the legal and other elements, which in the view of the Committee could be included in legally binding or non-legally binding instruments in the Council of Europe framework.

An eye will be kept on the Council of Europe activities, as there will be a need to ensure co-existence between the EU and Council of Europe frameworks as they will complement each other. It remains to be seen what recommendations from the CAHAI will be followed and what will be adopted under a legislative and non-legislative format.

As presented in this section, the regulation on AI media application is only in its infancy, and more will materialise in legally binding provisions in a near future.

12.2 Opportunities in the time of the pandemic

***Contributors**: Filareti Tsalakanidou (CERTH)*

The ongoing COVID-19 pandemic has abruptly disrupted our lives, has deeply affected society and has re-shaped the global economy. It has also undoubtedly boosted and in some cases catapulted AI research, with governments, researchers and industries trying to explore and exploit AI's potential, first and most in health and medicine but also in supply chain management, education, work, communication, etc.

The pandemic did not leave the media & entertainment industry unaffected. Consumer behaviours changed – in some cases overnight – accelerating changes that have been slowly manifesting during the last years. This in return resulted in significant shifts in the industry's internal dynamics (e.g. shifting from movie theatres to streaming services, renewed interest and trust in broadcast news, etc.) reshaping its future and overturning previous assumptions and expectations.

Obviously, the main factor driving the shifts in the media and entertainment industry was the abrupt migration towards digital consumption of media content. People could not go to the cinema or to concerts and live music events, they were reluctant to buy a print newspaper or book, and in most cases they were forced to stay at home for long periods of time with TV, podcasts, streaming services (for audio and film), online news sites, online games, e-books, and social media communities being their companion as they were trying to inform and entertain themselves or simply get through the day. This resulted on huge demand of online content consumption but not only that. It also fuelled e-commerce, with pandemic restrictions and humans' fear massively preventing people from going to the physical stores and forcing a lot of people to purchase goods online for the first time. This, on one hand, increased internet advertising of consumer products, benefitting the advertisement sector, but also increased ad revenue for other media companies whose channels were used for online ad placing.

[637] Council of Europe, CAHAI held its 6th and final plenary meeting, December 2021, https://www.coe.int/en/web/human-rights-rule-of-law/-/the-cahai-held-its-6th-and-final-plenary-meeting

Another trend accelerated by the pandemic was a generational shift towards younger audiences that increasingly set trends (e.g. by catapulting to fame TikTok personalities) and also a shift from producers to content creators[638] (while large production companies had to postpone or shut down production due to restrictions, social media were flooded by content created by independent creators).

The pandemic and the shift towards digital consumption also increased internet connections worldwide, with broadband connections showing a growth rate of 3.2% between 2019 and 2020, while smart phone connections have risen by 6.5%[638]. More connections, more devices and increased online activity all over the world resulted in an extraordinary growth of 30% in data consumption, with this growth trend projected to persist in the next few years.

All the aforementioned changes converge in a long developing trend: meet the users where they are, whether this is the internet, their home or the Metaverse, their PC or their mobile phone. And also give users what they want. The explosion of the online available media content has left users with a vast amount of choices and little time to spend on things they do not really like, making media companies enter an arms race to catch and most importantly retain user's interest and compete for the revenue generated from this interest.

Besides business or creative challenges, the pandemic also brought many media companies as well as the whole world face to face with unprecedented challenges concerning everyday work routines. Millions of professionals from the media and entertainment industry had to work from home, forcing managers to come up with emergency plans to re-invent how operations can continue to run smoothly in an abruptly changed reality.

But where AI stands with regard to all these pandemic-induced shifts in the media and entertainment sector? How can it help media companies amid new realities, new challenges and rapidly changing consumer behaviours? Based on the analysis of the previous sections of this deliverable, it is safe to predict that AI has an increasingly important role to play in the ever-changing media industry landscape. Below, we offer some insights on the role of AI in the media industry in the pandemic and post-pandemic era.

Users hungry for more content – How to facilitate new content generation? As mentioned above, the consumption of content skyrocketed during the pandemic, with users eager to spend their time at home as enjoyably as possible but also anxious to catch up with pandemic news and get trusted information. With content demand on the rise, the need to produce more content becomes pressing for media and entertainment companies. AI can play a double role in facilitating this trend: enable the automatic production of media content (from films and music to news) but also assist creators in the creative process, combining human creativity and machine intelligence and opening new roads for creative expression.

[638] PwC, Perspectives from the Global Entertainment & Media 2021: https://www.pwc.com/us/en/industries/tmt/library/assets/pwc-web-ready-pwc-outlook-perspectives-2021-2025.pdf

More and more content... – How to manage it? The pandemic saw a rise in the production of online media content both by the users but also by independent creators. At the same time, the audiences' demand for content they had not seen before led media companies, especially in the film/TV/music industry, to bring back old classics. The need to build large content libraries and exploit and make available all content calls for AI-driven content management functionalities. Advanced content indexing, search and retrieval will be required to allow journalists to mine the ever-growing media archives but also user generated content or to help streaming services to manage audio and video content. At the same time, AI-powered techniques for video and audio restoration or enhancement can help make old content more appealing to youngest audiences or provide new versions of old classics.

Tons of content everywhere – What to choose? Although, the pandemic has driven high rates of subscriptions for streaming giants like Netflix (with the trend seeming to continue well into the future) there is a limit to the number of subscriptions a house can make or the number of films a family can see or the number of books or news that an individual can read. With a plethora of offerings out there, media companies need to be able to attract user interest on their content and services by offering users exactly what they want and most importantly be able to continue to retain the user's interest, minimising churn. This is already unavoidably leading to a new phase of content delivery and experience, more focused on improving user experience and on retaining and creating increased value from already existing audience bases. AI is fundamental in this direction by analyzing user behaviour and real-word trends to offer users content that they will like or that it will be useful to them and by enabling highly personalised experiences but also new multi-platform experiences.

Users spending more and more time online – How to exploit and monetise the data they generate? An unquestionable effect of the pandemic is the shift to digital content consumption and the increased time spent online. As a result, users generate more data than ever with their online actions, behaviours and creations. This goldmine of data necessitates new innovative ways of analysis to fully exploit their potential. AI can help make sense of this data to improve user experience, content creation and delivery, to understand the audience, to predict future trends, to drive revenue growth and to plan in the short and long-term. To address, the ethical and legal concerns that such an analysis may justifiably raise, it is necessary to comply with relevant regulations such as the GDPR or the AI Act and adopt an ethics-by-design AI development approach.

Metaverse – How to be a part of it? With the pandemic forcing most of us to work at home and get entertained or communicate with friends and coworkers online, the Metaverse seems the next logical step regarding Internet evolution. AI has a big role to play not only in creating the Metaverse experience but also in facilitating media companies to move their business in this virtual world, by transforming existing content to fit the Metaverse, by allowing users to consume content in this world in novel multimodal ways, by designing sentient AI agents that will be our guides, assistants or companions (e.g. virtual characters that help us find content that we like, digital reporters providing the news of the day, or companions that play games with us), and by allowing part of the existing content creation, curation and delivery processes to take place in the Metaverse.

A sea of disinformation – How to navigate it and who to trust? The last couple of years have seen a sharp rise in online disinformation mainly related to the pandemic and vaccines. As a result, journalists are being overwhelmed by the gigantic volume of fake news, struggling to keep up with fact-checking processes, and in many cases failing, especially under the pressure to get the scoop first and beat the competition in reporting it, in a 24-hour news cycle that never slows down[639]. At the same time, the global audience, anxious for the present and the future, overwhelmed by the plethora of news and news outlets, and worried about disinformation, is thirsty for real news, facts, and news organisations and journalists that they can trust. AI can help journalists and the audience to navigate the Wild West of disinformation by offering fact-checking and content verification tools for newsrooms & content moderators but also for the public while also providing the tools to study the disinformation phenomenon in depth (how disinformation starts, how it spreads, who are the main disseminators, how it affects people, etc.). The pandemic was also shown to be a real opportunity for traditional news outlets, seeing consumption of TV news and print press (the online version of it) increasing considerably, reflecting the audience's need for accurate and reliable news sources that they cannot find on like social media[640]. While trust in news has been globally declining over the last decade, it has actually grown by six percent in 2021 in the wake of the pandemic, while the trust gap between the news sources people generally rely on and the news they find in social media and search further increased, climbing to 20% (44% vs. 24%)[640]. Traditional media have the opportunity to consolidate their position as reliable news sources against social media and further consolidate audience trust by widely adopting AI-powered tools for rigorous fact-checking and content verification.

Workplace disruption – How to deal with it? As with all other aspects of personal and professional life, the media industry faced major disruptions in the wake of the pandemic that upended established workflows and threatened the smooth operation of organisations. With the pandemic still looming over us, CEOs, business managers and data officers need to ensure that employees are protected, on one hand, and keep the business running as usual, on the other. This balance act becomes increasingly challenging, requiring new operational models that will adapt to changing conditions, effectively combining traditional workplace workflows with remote work. In the quest for flexibility, resilience and agility, AI can facilitate the implementation of new, more flexible operational models and task workflows, by automating existing processes and tasks but also by providing the tools (e.g. AI assistants) that will allow media professionals to seamlessly do the work independently of where they are or what resources are available.

[639] M. Scott, Politico, 'It's overwhelming': On the frontline to combat coronavirus 'fake news' (2020): https://www.politico.eu/article/coronavirus-fake-news-fact-checkers-google-facebook-germany-spain-bosnia-brazil-united-states/

[640] N. Newman et al., Reuters Institute Digital News Report 2021 (2021): https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2021-06/Digital_News_Report_2021_FINAL.pdf

More users, more devices, more services, more content, more data, more demand – What about processing power? The pandemic has brought an explosion in content and data created and consumed online as well a sharp increase in demand for new media services available to the user any time and any place. There is an increasingly growing need for processing power that will support demand and facilitate the AI-powered applications for the analysis of gigantic volumes of data. Emerging technologies like quantum computing for AI and AI chips can be a game changer in this direction.

In a few words, the pandemic created exciting new opportunities for the growth of the media and entertainment industry and the expansion of their services and audience, while at the same time brought significant operational and creative challenges. These trends consolidate the position of AI as a transformative power in this industry, capable of revolutionising how operations run and how content is created, delivered and consumed. To exploit this potential, media companies should define clear strategies for operationalising AI across the whole industry, pay close attention to new AI advances and breakthroughs being ready to leverage them to improve business, and invest on AI training of their personnel and collaborations with research/academia to develop new exciting AI-powered applications for the media sector. There is overwhelming agreement that AI will drive the majority of innovation across nearly every industry sector in the next decade, especially in the wake of the pandemic, as was shown by a recent IEEE survey[641]. 51% of technology leaders surveyed said that due to the pandemic their companies accelerated the adoption AI and machine learning (second only to cloud computing at 60%) while 66% strongly agrees that AI will drive innovation.

12.3 Climate crisis and AI

Contributors: *Filareti Tsalakanidou (CERTH)*

According to the United Nations, "*climate change refers to long-term shifts in temperatures and weather patterns. These shifts may be natural, such as through variations in the solar cycle. But since the 1800s, human activities have been the main driver of climate change, primarily due to burning fossil fuels like coal, oil and gas. Burning fossil fuels generates greenhouse gas emissions that act like a blanket wrapped around the Earth, trapping the sun's heat and raising temperatures. […] Energy, industry, transport, buildings, agriculture and land use are among the main emitters.*"[642] Greenhouse emissions are at their highest levels in 2 million years and continue to rise, resulting in Earth's temperature rising about 1.1°C since the 1800s. According to the same UN article, the consequences of this rise are enormous and dire: intense droughts, water scarcity, severe fires, rising sea levels, flooding, melting polar ice, catastrophic storms and declining biodiversity that can affect our health, safety, housing, ability to grow food. People in many places around the world, especially in developing countries, are already facing the disastrous effects of climate change and experts expect that the number of climate refugees will be on the rise in the coming decades.

[641] IEEE, The impact of Tech in 2022 and beyond, https://transmitter.ieee.org/impact-of-technology-2022/
[642] United Nations, What Is Climate Change?: https://www.un.org/en/climatechange/what-is-climate-change

Climate change is forcing governments, industries and citizens to adapt and take action before it is too late, going into emergency mode against climate crisis. Global frameworks and agreements have been set in place to guide progress in cutting emissions, adopting alternative solutions and adapting to change, including the Paris Agreement[643], UN's Sustainable Development Goals[644], the UN Framework Convention on Climate Change[645] and the Paris Agreement.

The media and entertainment industry are not cut out of this reality. In fact, media is part of the problem since it consists of physical infrastructure (data centers, content delivery networks, access networks, etc.) that obviously require gigantic amounts of energy to operate. Some interesting statistics are presented below:

- The carbon footprint of the Internet (incl. infrastructure, our devices and supporting services) accounts for 3.7% of global greenhouse emissions, similarly to those of the airline industry, and is expected to double by 2025. Considering that Netflix and YouTube combined represent more than 50% of Internet traffic at peak times in North America, audio and video streaming make more than 60% of the global Internet traffic, and online gaming increases by 19% per year, it is obvious that the media industry contributes significantly to the problem[646].
- A report[647] published by Carbon Trust in 2021 shows that the media and entertainment industry accounts for about 46% of the carbon footprint of ICT[648]. The same report estimates that average carbon emissions per hour of video streaming in Europe for 2020 were 56g CO2 e per hour of video streaming while the equivalent energy consumption is 188Wh per hour of streaming.
- Netflix reported that fans spent more than 6 billion hours watching the top 10 shows in the first 28 days after each show was released in 2021. Based on the estimation above for the average carbon emissions per hour of video streaming, this translates to 1.8bn km of travel in a car – the current distance between Earth and Saturn[649].

---

643 The Paris Agreement: https://www.un.org/en/climatechange/paris-agreement

644 United Nations' Sustainable Development Goals: https://www.un.org/sustainabledevelopment/sustainable-development-goals/

645 United Nations' Framework Convention on Climate Change: https://unfccc.int/process-and-meetings/the-convention/what-is-the-united-nations-framework-convention-on-climate-change

646 Climate Care, Infographic: The Carbon Footprint of the Internet (2021): https://www.climatecare.org/resources/news/infographic-carbon-footprint-internet/

647 Carbon Trust, Carbon impact of video streaming: https://www.carbontrust.com/resources/carbon-impact-of-video-streaming

648 The following narrow definition of the carbon footprint of the media & entertainment industry is adopted: Media & entertainment comprises all electronic equipment utilised for media and entertainment purposes, including: TVs, cameras, and other E&M consumer electronics, as well as physical paper media and printing. The definition excludes cinemas, theatres, and other arenas or physical site events (e.g., sports), and content creation such as film and TV production.

649 M. Sweney, Guardian, Streaming's dirty secret: how viewing Netflix top 10 creates vast quantity of CO2 (2021): https://www.theguardian.com/tv-and-radio/2021/oct/29/streamings-dirty-secret-how-viewing-netflix-top-10-creates-vast-quantity-of-co2

- A report by the Shift Project[650], estimates that in 2018 online video viewing generated more than 300 Mt CO_2, i.e. 1% of global emissions or as much as Spain emits, while video on demand services by streaming giants produced more than 100 Mt CO_2e/year or 0.3% of world emissions. Given that since 2018 online video viewing and streaming have witnessed an unprecedented growth, also fuelled by the pandemic, we can safely assume that greenhouse gas emissions have increased even more since then.
- When a global brand displays 40 billion 200kB ad impressions, this can generate up to 8,000 tons of CO_2[651]. A typical online ad campaign would emit 5.4 tonnes of CO_2, which is the equivalent of 43% of the average annual carbon footprint of a person in the United Kingdom[652].
- Movies with a budget of $50 million dollars typically produce the equivalent of around 4,000 Mt of CO_2[653]. Blockbuster films' average carbon footprint is 3,370 Mt (or about 33 Mt per shooting day) while one-hour scripted dramas for TV had 77 Mt of CO_2 emissions per episode[654].

There are already a lot of new and promising initiatives that can help the media industry calculate, manage and reduce carbon emissions, like Albert[655], Ad Net Zero[656], the AdGreen calculator[657], and the IPA carbon calculator[658].

There are a lot of ways in which AI could help the media industry reduce its carbon footprint. For example, automation of film and music production processes and workflows that traditionally require human presence could reduce direct emissions such as those from fuel used by production vehicles and generators but also indirect emissions from travel and accommodations[654]. Similarly, automatic generation of new content (news summaries, visuals, game assets, movie trailers, etc.) reduces significantly the time and work effort required to produce the content and thus related emissions from human professional activity, buildings, equipment, commute etc. Targeted advertisement that knows when and where to find the user could reduce the number of online advertisements and thus relevant traffic, which nowadays is far from negligible. AI at the edge running on the user's or professional's device can significantly

---

650 The Shift Project, Climate crisis: the unsustainable use of online video – The practical case for digital sobriety (2019): https://theshiftproject.org/wp-content/uploads/2019/07/2019-02.pdf

651 J. Benon, The Drum, How the media industry is in a unique position to reduce the internet's CO2 impact (2022): https://www.thedrum.com/industryinsights/2022/01/05/how-the-media-industry-unique-position-reduce-the-internet-s-co2-impact

652 O. Oakes, Mediatel Nees, Carbon footprint of digital ads laid bare by Good-Loop tool (2021): https://mediatel.co.uk/news/2021/06/03/carbon-footprint-of-digital-ads-laid-bare-by-good-loop-tool/

653 Columbia Climate School, Cut! How the Entertainment Industry is Reducing Environmental Impacts (2018): https://news.climate.columbia.edu/2018/03/29/entertainment-industry-sustainability/

654 T. Spangler, Variety, Hollywood Studios Release Carbon-Emissions Report, Showing Wide Variance Among Productions (2021): https://variety.com/2021/film/news/sustainable-production-alliance-carbon-emissions-report-1234942580/

655 Albert: https://wearealbert.org/

656 Ad Net Zero: https://adassoc.org.uk/ad-net-zero/

657 The AdGreen Carbon Calculator: https://weareadgreen.org/carbon-calculator

658 IPA Media Carbon Calculator: https://ipamediaclimatecharter.co.uk/media-carbon-calculator/how-to-use-the-calculator/

reduce processing power and reduce the volume of data being sent across networks while quantum computing is set to go even further. And AI can help the media and entertainment industry calculate and predict carbon footprint of its different processes and optimise them, thus enabling resource management focused on sustainability.

Since AI is making everything faster and more efficient, it is not unreasonable to expect that it can help reduce the industry's environmental footprint. But that assumption oversees a very basic fact. AI is not something intangible that somehow just happens in a void. It requires huge computational power and physical resources to be able to deliver all the amazing applications discussed in the previous sections, from recommender systems, to smart visual search, to automatic content creation, to personalised services. AI is by definition a data hungry technology that requires enormous amounts of data to train models that exhibit high accuracy in different tasks like image classification, natural language processing, or prediction analytics.

To grasp the problem let us consider a simple example, presented in the essay "*Anatomy of an AI system*"[659]: for Amazon's Alexa to be able to answer a question, play a song or switch on a device, continuous learning from the interactions between Alexa and the user is required; the computational resources for this however are much higher than the energy and labor it would take for a human to flick a switch.

Training AI models requires large amounts of compute resources. In 2018, OpenAI released an analysis[660] showing that since 2012, the year deep neural networks were introduced, the amount of compute used in AI training has been increasing exponentially with a doubling time of 3.4 months (compared to the two years of Moore's law), increasing by an astounding 300,000 times by 2018 (see Figure 113). An even sharper rise was observed in training NLP models, especially with the development of BERTs. As discussed in a recent report by Google and Berkeley researchers[661], OpenAI's GPT-3 language model (trained on almost 500 billion words and using 175 billion parameters) produced the equivalent of 552 metric tons of CO_2 during its training, which is the equivalent of driving 120 passenger cars for a year[662] or driving to the moon and back. Google's chatbot Meena consumed 96 metric tons of CO_2e, which is almost the same as powering 17 houses for a year.

[659] K. Crawford and V. Joler, "Anatomy of an AI System: The Amazon Echo As An Anatomical Map of Human Labor, Data and Planetary Resources," AI Now Institute and Share Lab (2018): https://anatomyof.ai
[660] OpenAI, AI and Compute (2018): https://openai.com/blog/ai-and-compute/
[661] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean, Carbon Emissions and Large Neural Network Training (2021): https://arxiv.org/abs/2104.10350
[662] J. Kahn, A.I.'s carbon footprint is big, but easy to reduce, Google researchers say (2021): https://fortune.com/2021/04/21/ai-carbon-footprint-reduce-environmental-impact-of-tech-google-research-study/

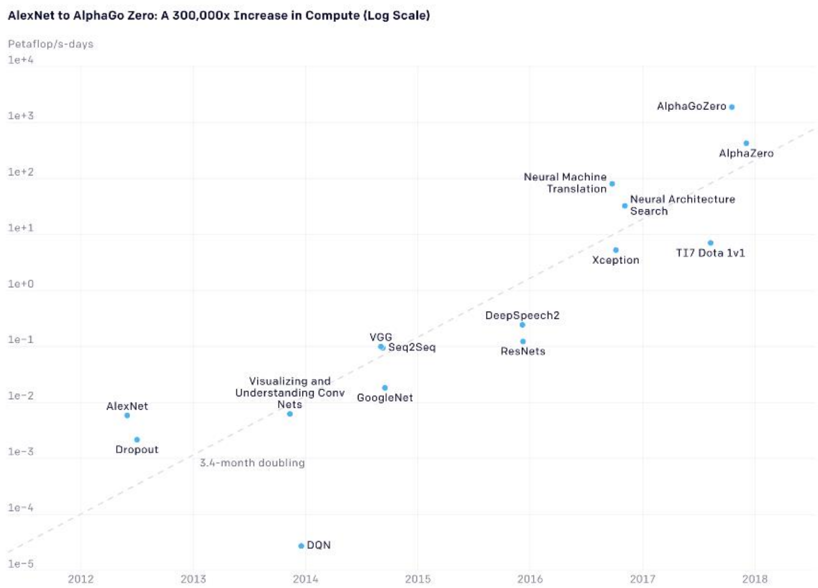


Figure 113: Increase in AI training compute from 2012 to 2018[663].

The same study identified several factors that contribute to the carbon footprint of AI training: i) algorithm improvement (e.g. optimised architectures that require less rounds of training and less data); ii) processor improvement (e.g. building AI chips that accelerate model training and serving, instead of using GPUs); iii) datacenter improvement (e.g. getting access to energy-optimised datacenters, powered by green energy).

A recent paper examines the environmental impact of AI from a holistic perspective, presenting challenges and opportunities for sustainable AI computing the ML development process — Data, Experimentation, Training, and Inference — for a variety of AI use cases at Facebook, including recommendation, computer vision and NLP. The study shows that the AI algorithm used affects significantly the carbon footprint: for example, training the Switch Transformer that has 1.5 trillion parameters produces significantly less carbon emission than GPT-3 (750 billion parameters), highlighting the environmental advantage of network architectures. The success of an AI system should be measured in terms of its overall impact. This will help us select the correct levels of complexity for a particular problem, which can lead to a significant reduction in energy consumption, cost and time[664].

Data centers are also an important part of the equation. Used for the storage and processing of the data required to train AI algorithms, data centers comprise millions of machines that consume huge amounts of energy. To understand the size of their environmental impact, with the current growth rate Ireland's data center sector is projected to consume about 23% of the

[663] Image source: OpenAI - https://openai.com/blog/ai-and-compute/
[664] B. Mullins, Time to tackle AI's impact on the environment (2021): https://sifted.eu/articles/ai-environmental-impact/c

country's total energy demand by 2030, forcing the Irish Government to consider placing restrictions on data center building or even banning it, in order to meet targets for emissions and renewable energy[665]. A data center's efficiency and regional placement may have a major impact on emissions. For example, a server in Quebec (which is dominated by low-carbon hydroelectricity) may emit as many as 35 times less CO_2 per KWh as a server in Iowa (where, after wind energy, coal is the most common electricity source)[666].

The increase in AI training compute was mainly motivated by the need for more accurate models. A new term has been introduced by Schwartz et al.[667] to describe the phenomenon: ***Red AI***. Red AI is AI seeking to obtain state-of-the-art results in accuracy through the use of massive computational power. On contrast, ***green AI*** aims to be more environmentally friendly, interested not only in better models but also more sustainable ones, seeking to achieve great performance without increasing, or even decreasing, computational cost.

To truly benefit from AI's great potential, we need a better understanding of AI's growing carbon footprint so that we are able to find the balance between transformative applications for media professionals and users, on one hand, and impact of those applications in the deteriorating climate crisis, on the other.

Leaders in the media industry must be proactive in making sure that the AI technologies they adopt take environmental considerations into account while AI researchers and developers should make sure that a green AI approach is adopted to address these concerns. An understanding of the costs of building and deploying AI models not only financially but also in terms of environmental impact must be developed across the board[668]. The mentality that more data or bigger models is always better should be reexamined while technologies such as evolutionary learning that adapt a model based on new information rather than retraining it from scratch can help reduce computational costs.

To this end, environmental standards should be developed to ensure the mitigation of environmental impacts while green AI certifications could be introduced to promote green AI development by AI researchers and tech companies. At the same time, media companies deploying AI technologies should develop industrial guidelines promoting the procurement of green AI[669]. Ensuring that AI for media is developed not only transparently and responsibly but also sustainably, the media and entertainment industry can reap the benefits of AI, without its adoption being counterproductive in the fight against climate change.

---

[665] P. Judge, Data Centre Dynamics, Irish government could restrict data center building (2021): https://www.datacenterdynamics.com/en/news/irish-government-could-restrict-data-center-building/

[666] F. Rice, Greening AI: Rebooting the environmental harms of machine learning (2021): https://www.corporateknights.com/clean-technology/greening-ai/

[667] R. Schwartz, J. Dodge, N. Smith, and O. Etzioni, Green AI (2109): https://arxiv.org/pdf/1907.10597.pdf

[668] B. Mullins, Time to tackle AI's impact on the environment (2021): https://sifted.eu/articles/ai-environmental-impact/

[669] P. Dhar, The carbon impact of artificial intelligence (2020), https://www.nature.com/articles/s42256-020-0219-9

13 Web version of the Roadmap

The current report provides a detailed overview of the current state of play and future research trends with regard to media AI, covering various aspects analysed in detail in sections 3-12 and resulting in a comprehensive document of 390 pages. In addition to the full report, and in order to help potential readers focus on the aspects that are most relevant to their own expertise and interests, we have also created **a web version of the report** as part of the AI4Media website. This allows readers to easily go through the contents of this report though a user-friendly visual interface and only read/download the subsections covering the topics that interest them.

The web version of the report is available at https://www.ai4media.eu/roadmap-ai-for-media/. Figure 114 presents the visual interface. As can be seen, the report has been divided in different parts, roughly corresponding to its subsections, which can be directly downloaded as separate pdf files (of only a few pages) or viewed online by simply clicking on the relevant button (see Figure 114(b)).

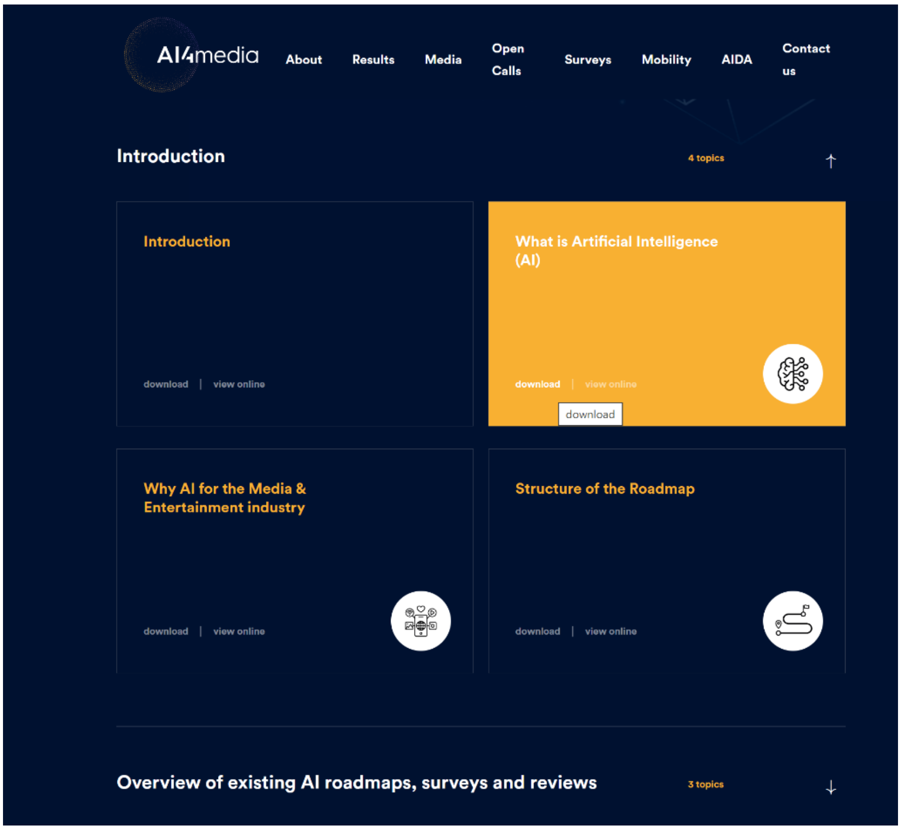


Figure 114: The web version of the AI4Media Roadmap on AI for the Media Industry (D2.3). The user can download or view online the parts of the report they are interested in by simply clicking on the corresponding buttons.

The structure of the report as roughly implemented in the website can be seen in Figure 115 and Figure 116.

The web version of the report will be fully operational at the end of March 2022. We believe that this web version will greatly facilitate the wide dissemination of the AI4Media Roadmap, since it will allow interested readers from the AI research community, the media industry, relevant public or civil society organisations, and the general public to easily locate and read the sections that are of interest to them.

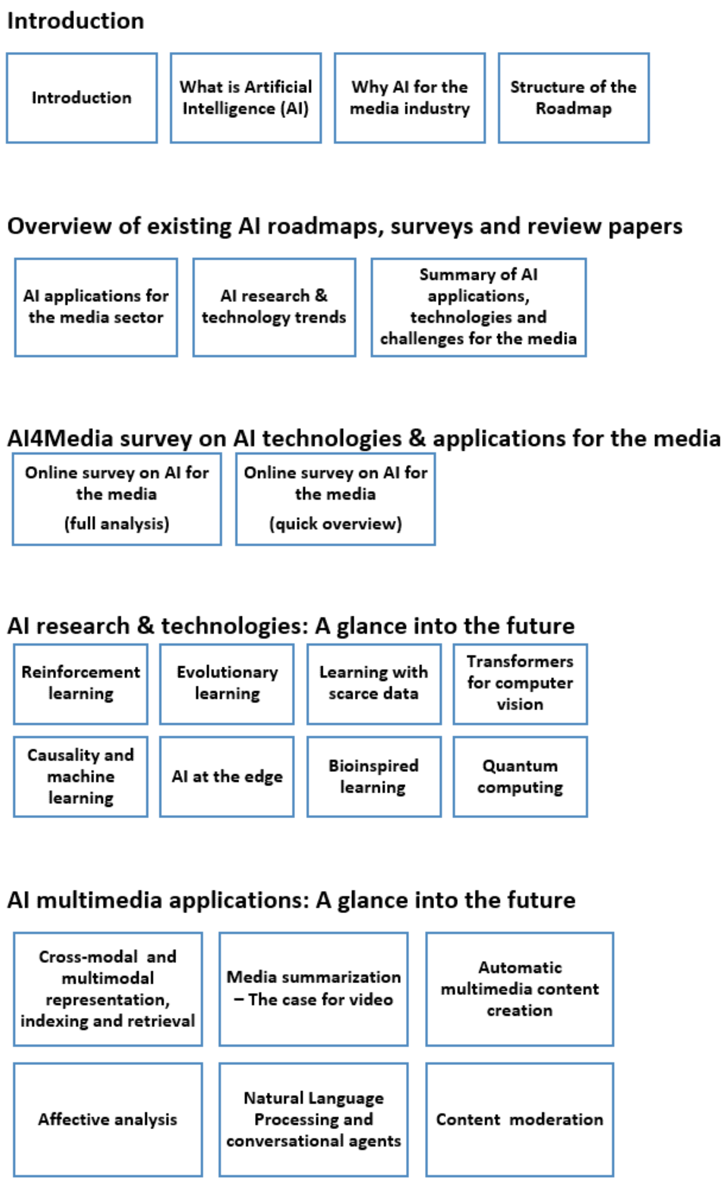


Figure 115: The (abstract) structure of the web version of the AI4Media Roadmap on AI for the Media Industry (part 1).

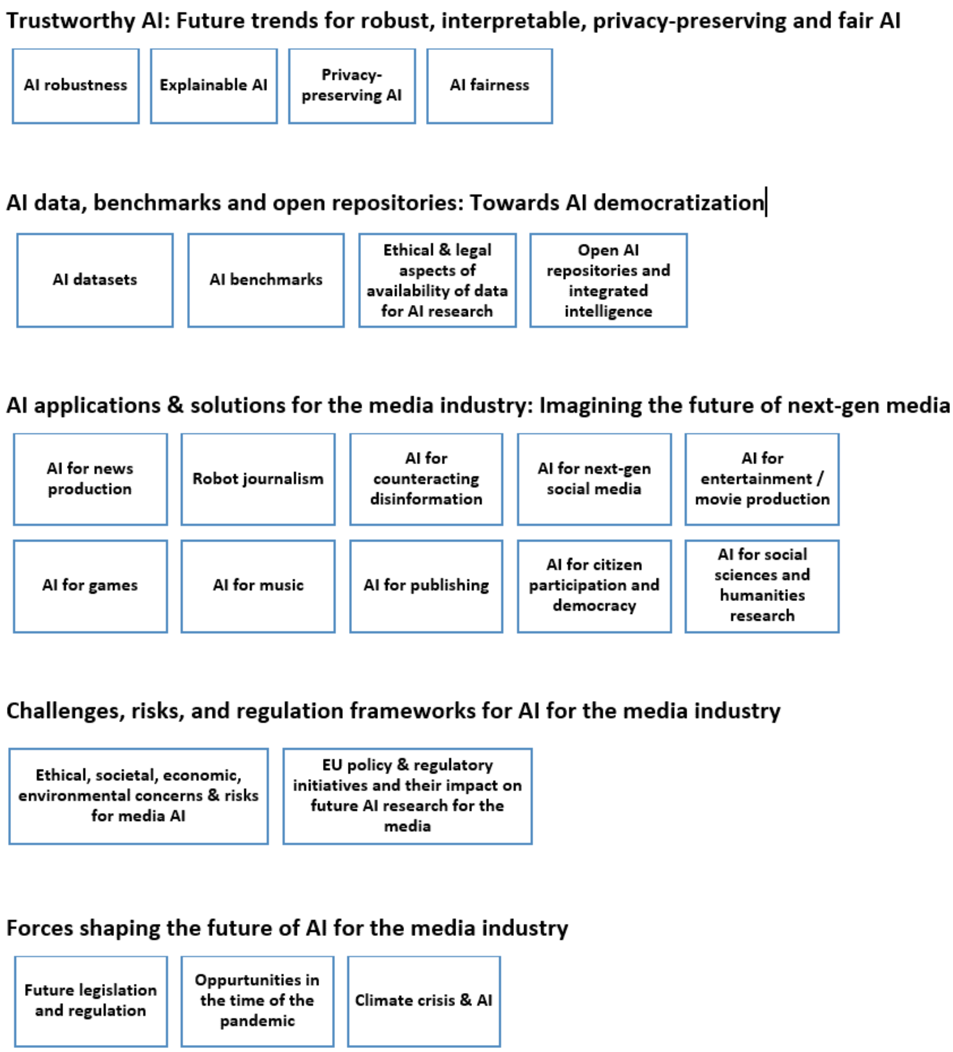


Figure 116: The (abstract) structure of the web version of the AI4Media Roadmap on AI for the Media Industry (part 2).

14 Conclusions

Based on a multi-disciplinary state-of-the-art analysis, a public survey targeted at AI researchers and media professionals, and short white papers that focused on different AI technologies, AI applications, and media sectors, this deliverable offers a detailed overview of the complex landscape of AI for the media industry.

The ***state of the art analysis*** provides a clear picture of the most important applications of AI in the media and entertainment industry, including AI assistants, smart recommender systems, content personalisation, automatic content synthesis, multi-modal content search, multi-lingual translation, audience analysis, social media trend detection, and forecasting. Additionally, it identifies technology trends such as reinforcement learning, evolutionary learning, learning with scarce data, transformers, causal AI, AI at the edge, bioinspired learning and quantum computing, highlighting their potential to revolutionise multimedia analysis applications like the aforementioned ones.

The ***survey analysis*** offers insights on what people developing AI technologies or working in the media industry think with regard to media AI. It highlights where the most important benefits are expected (automation & optimisation of routine tasks, personalisation of content and services, increased operational efficiency, enhancement of current services) but also what are the main risks (unethical use of AI, bias, lack of explainability, failure in critical missions, high expectations but low return). In addition, it sheds light to existing practices with regard to AI strategies, ethical guidelines, and AI skills in media and research organisations.

The ***white papers on selected AI technologies and applications*** as well as on issues of ***trustworthy AI*** offer a clear overview of the current status of each technology, the drivers and challenges for its development and adoption, future trends with regard to their deployment in the media, and goals for next decades. They also provide imaginative scenarios, showcasing how these AI innovations could help transform the media industry in the future, by providing reliable solutions to its most pressing problems: increasing demand for new content, automatic analysis of audio-visual content, language barriers, lack of data to train algorithms for new tasks or domains, and need for AI that can be trusted.

In addition, the ***discussion about AI data, benchmarks, open repositories and integrated intelligence*** provides useful insights on relevant research challenges involved in the development of AI algorithms and software, highlighting the need for sharing data and software to ensure democratised access to AI but also the need to respect data privacy and follow ethical guidelines with regard to data collection and processing.

The ***white papers focusing on how AI can be used in different sectors of the media industry***, including news, social media, games, music, film, publishing, etc. reveal a landscape where the opportunities for the use of AI are enormous while the variety of tasks across the media supply chain that AI can improve, assist, automate, expand or create is limitless. AI can have a truly transformative influence on the media sector, reinventing the business model of media organisations, establishing new ways of work and increasing the productivity and creativity of the workforce, and finally transforming and enhancing the user experience across platforms.

Besides the technological overview & application examples, the report offers a much needed ***insight on the social and ethical implications of AI for media***, identifying the most prevalent societal concerns and risks: bias and discrimination; media (in)dependence and commercialisation; inequality to access to AI; labour displacement, monitoring and profession transformations; privacy, transparency, accountability and liability; manipulation and misinformation as an institutional threat; and environmental impacts of AI. In addition, it identifies measures that could help counteract these risks.

Similarly, the ***analysis of proposed EU policy initiatives*** like the AI Act or the DSA and the corresponding commentary sheds light on the legislative landscape, focusing on the impact of the proposed legislation on AI research and media organisations and highlighting the trend for more regulation with regard to media AI.

Finally, the ***discussion about the impact of regulations, the pandemic and climate crisis*** on AI for the media highlights great opportunities but also great risks and points out the need for close monitoring of legislative initiatives; for effectively operationalising AI across the whole industry; and for the development of environmental standards and green AI certifications.

The different parts of the deliverable highlight the role of AI as a potentially transformative power in the media industry, capable of revolutionising how operations run and how content is created, delivered and consumed. To exploit this potential, media companies should define clear strategies for operationalising AI across the whole industry, pay close attention to new AI advances and breakthroughs being ready to leverage them to improve business, and invest on AI training of their personnel and collaborations with research/academia to develop new exciting AI-powered applications for the media sector. In addition, they should focus more on issues of trust, ethics and accountability and enforce strong ethical frameworks to closely monitor and effectively mitigate the many societal, economic, environmental and ethical risks of media AI.

There is overwhelming agreement that AI will drive the majority of innovation across nearly every industry sector in the next decade. The media industry should be ready to exploit new AI advances but also mitigate possible risks, in order to enjoy the full potential of this technology.

15 Appendix

This Appendix presents the questionnaires used in the two AI4Media online surveys analysed in section 4: i) AI4Media survey on AI for the Media Sector (version for AI community) in section 15.1; ii) AI4Media survey on AI for the Media Sector (version for media professionals) in section 15.2; and iii) AI4Media survey on Media AI in the service of Society & Democracy in section 15.3.

15.1 AI4Media survey on AI for the Media Sector (version for AI community)

AI4Media online survey on AI for the Media Sector

For AI researchers and software developers

About this survey

About this survey

This questionnaire has been created in the context of the **AI4Media H2020 research project** (https://www.ai4media.eu/) and aims to help the AI4Media consortium in the development of a **Roadmap on Artificial Intelligence technologies and applications for the Media Sector**. Our aim is to collect the opinions of the AI research community with regard to the most important trends, benefits, challenges, risks, and facilitators for the use of AI in the media sector. A similar questionnaire will also be shared with representatives of the broader media industry (news industry, film/TV/radio industry, gaming industry, etc.). The results of this survey will be published as part of AI4Media's public deliverable "*D2.3 – AI technologies and applications in media: State of Play, Foresight, and Research Directions*".

The survey is **anonymous**, so there is no need to give your name or the name of your organisation. Data is not linked to respondents and is only going to be used in an aggregated way in this deliverable or any subsequent AI4Media publications.

Survey structure

The survey includes the following sections: (1) Professional backgound of survey respondents; (2) Future AI technology trends for the media sector; (3) AI benefits for the media sector; (4) AI risks for the media sector; (5) AI challenges for the media sector; (6) Adoption of AI solutions by media sector; (7) AI ethics & AI regulation.

Each section usually includes 1-2 multiple choice questions + complementary open-ended questions (allowing you to elaborate, in case the previous choices were not enough). The questionnaire takes around **10-15 minutes** to complete.

Some survey questions and answers have been inspired by Deloitte's survey on "State of AI in the Enterprise", 2nd edition (https://tinyurl.com/2p8kzptv).

About AI4Media

Motivated by the challenges, risks and opportunities that the widespread use of AI has brought to the media, society and politics, the EU-funded AI4Media project aspires to establish a centre of excellence and a wide network of AI researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the Media sector, to make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and to reimagine AI as a beneficial enabling technology in the service of Society and Media. The project also includes a funding framework through open calls, an AI Doctoral Academy, and seven use cases to demonstrate the impact on the media sector.

More information about AI4Media can be found at https://www.ai4media.eu/. You can also watch our 3-min concept video on YouTube (https://tinyurl.com/2p9yrw3f).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911.

Personal data collection

No personal data of the survey respondents will be collected or stored. Answers to this survey will only be used in an aggregated way for the purposes described above.

Contact info

In case of questions about this survey, please send an email to filareti (at) iti (dot) gr

Consent

By clicking on **SUBMIT** after you have completed the last section of this questionnaire, you confirm that the information you have received has been sufficient and that you agree with the way this survey is being conducted.

Thank you for your participating in the survey! Your feedback is important.

Next

Page 1 of 8

* Required

Professional background

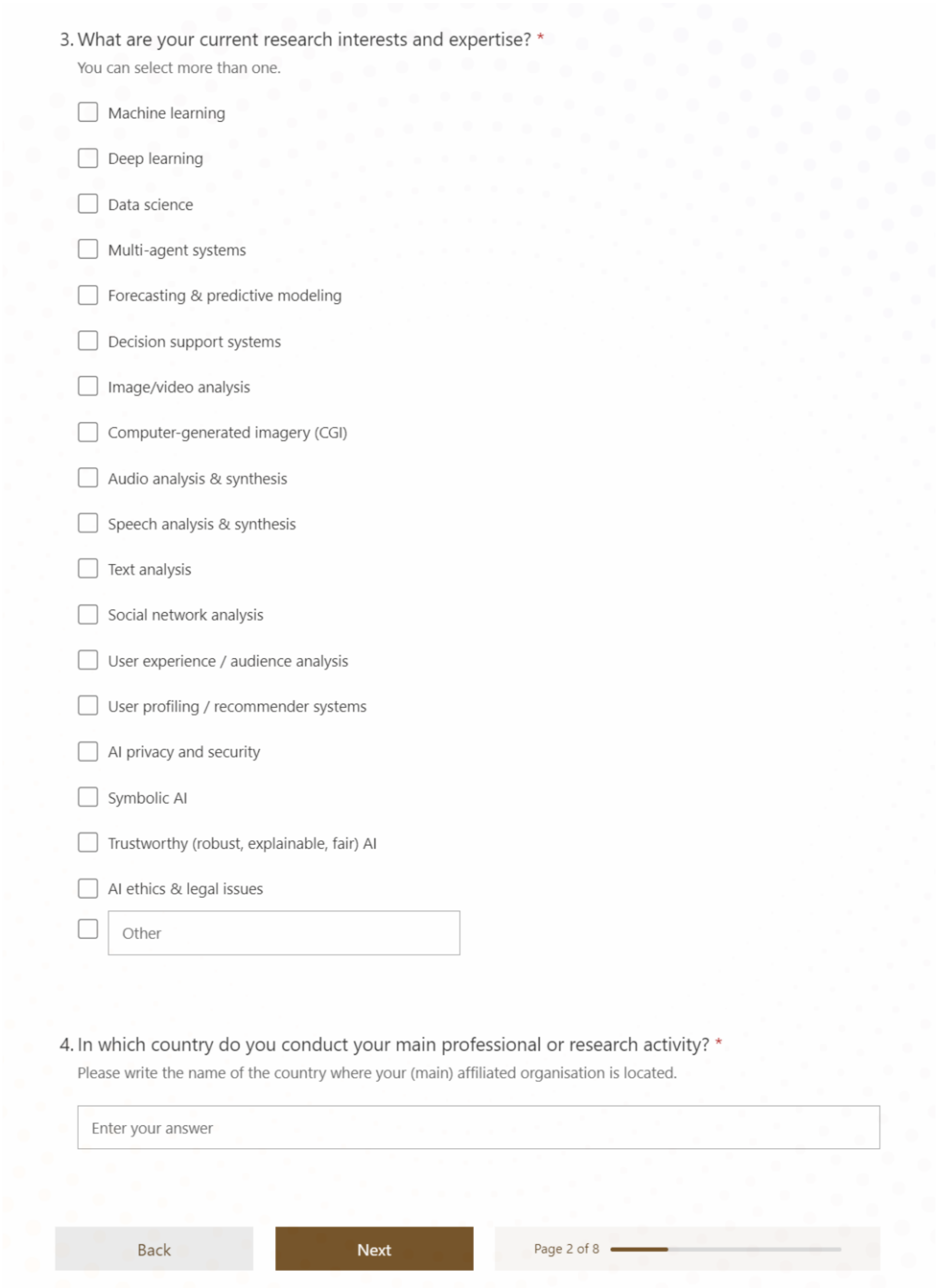
Professional backgound of survey respondents

1. What type of organisation do you work for? *

- (•) Academic institution
- ( ) Research institution
- ( ) Public Service Media (PSM)
- ( ) Media & entertainment industry
- ( ) ICT industry
- ( ) Non-governmental organization (NGO)
- ( ) Independent authority
- ( ) Public sector
- ( ) Other

2. What kind of position do you hold in your organisation? *

- ( ) AI researcher
- ( ) AI software developer
- ( ) Data scientist
- ( ) AI ethics/legal expert
- ( ) Innovation or R&D manager
- ( ) Other



3. What are your current research interests and expertise? *

You can select more than one.

- [ ] Machine learning
- [ ] Deep learning
- [ ] Data science
- [ ] Multi-agent systems
- [ ] Forecasting & predictive modeling
- [ ] Decision support systems
- [ ] Image/video analysis
- [ ] Computer-generated imagery (CGI)
- [ ] Audio analysis & synthesis
- [ ] Speech analysis & synthesis
- [ ] Text analysis
- [ ] Social network analysis
- [ ] User experience / audience analysis
- [ ] User profiling / recommender systems
- [ ] AI privacy and security
- [ ] Symbolic AI
- [ ] Trustworthy (robust, explainable, fair) AI
- [ ] AI ethics & legal issues
- [ ] Other

4. In which country do you conduct your main professional or research activity? *

Please write the name of the country where your (main) affiliated organisation is located.

Enter your answer

Back | Next | Page 2 of 8

* Required

Future AI technology trends for the media sector

In this section, we explore future AI technology trends for the media sector.

5. In your opinion, what is the potential of the following AI technology trends to significantly help or transform (parts of) the media sector in the next decade?
*

Only to be answered by those who understand these technologies and their potential. Please assess the potential of each technology on a scale from 1 (limited potential) to 5 (significant potential).

	1	2	3	4	5	Don't know
Learning with limited data	○	○	○	○	○	○
Reinforcement learning	○	○	○	○	○	○
Evolutionary learning	○	○	○	○	○	○
Transformers for computer vision	○	○	○	○	○	○
Bioinspired learning	○	○	○	○	○	○
Causal AI	○	○	○	○	○	○
Emotion AI	○	○	○	○	○	○
Multi-language NLP	○	○	○	○	○	○
Automatic content analysis & knowledge extraction	○	○	○	○	○	○
Automatic content creation	○	○	○	○	○	○
Explainable AI	○	○	○	○	○	○
Trusted & fair AI	○	○	○	○	○	○
AI at the edge	○	○	○	○	○	○
Quantum computing	○	○	○	○	○	○

6. Are there other important AI technology trends that could help the media sector, which are not mentioned in the previous question?
This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back | Next | Page 3 of 8

AI benefits for the media sector

In this section, we explore the benefits that AI technology can bring to the media sector.

7. What are the main benefits that AI can bring to the media sector in the next decade?
*

Only to be answered by those who understand these concepts/benefits. Please assess the importance of each benefit on a scale from 1 (not important) to 5 (very important).

	1	2	3	4	5	Don't know
Personalisation of content & services	○	○	○	○	○	○
Automation & optimisation of routine tasks	○	○	○	○	○	○
Enhance current services & products	○	○	○	○	○	○
Create or enable new services, products, content	○	○	○	○	○	○
Enhanced decision making	○	○	○	○	○	○
Increased productivity & operational efficiency	○	○	○	○	○	○
Increased creativity	○	○	○	○	○	○
Big data analysis to identify business opportunities, maximise revenue or conduct journalistic research	○	○	○	○	○	○
Facilitate expansion to new markets / target new audiences	○	○	○	○	○	○
Better monetisation of collected knowledge & data (e.g. video archives)	○	○	○	○	○	○

8. Are there other benefits that AI can bring to the media sector in the next decade that are not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back | Next | Page 4 of 8

AI risks for the media sector

In this section, we explore the potential risks of AI technologies for the media sector.

9. What are you most worried about when it comes to the adoption of AI by the media sector? Both as a citizen as well as an AI researcher.
*

Please assess the following risks on a scale from 1 (not worried) to 5 (very worried).

	1	2	3	4	5	Don't know
Automation that shrinks human capital	○	○	○	○	○	○
Automation that limits human creativity	○	○	○	○	○	○
Unethical use of AI with destructive impact on individuals and society	○	○	○	○	○	○
Increased reliance on automated AI decisions without human in the loop	○	○	○	○	○	○
AI being a black box with little explanation on how decisions are made	○	○	○	○	○	○
AI models & software being susceptible to attacks	○	○	○	○	○	○
AI models being biased against groups of people	○	○	○	○	○	○
Inability to distinguish between created and real content	○	○	○	○	○	○
Exploiting user profiling on different platforms to create super profiles	○	○	○	○	○	○
Government not being able to effectively regulate AI	○	○	○	○	○	○
AI increasing the power of media tech giants, leading to monopolies	○	○	○	○	○	○
AI further amplifying echo chambers	○	○	○	○	○	○
Exploitation of AI to silence different voices in the media	○	○	○	○	○	○
Exploitation of media AI to hinder fundamental human rights, like freedom of expression	○	○	○	○	○	○

10. Are there other potential risks from the adoption of AI by the media sector, which are not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Here you can also elaborate on risks that are specific to your own research area.

Enter your answer

Back Next Page 5 of 8

AI challenges for the media sector

In this section, we explore the challenges of developing AI technology for the media sector.

11. What are the most important challenges AI researchers face when developing AI algorithms and tools for the media sector?
*

Only to be answered by those who understand these concepts/challenges. Please assess each challenge on a scale from 1 (not important) to 5 (very important).

	1	2	3	4	5	Don't know
Lack of real-world data to train/test AI (e.g. in different languages)	○	○	○	○	○	○
Media industry not sharing their data	○	○	○	○	○	○
Lack of effective AI benchmarks	○	○	○	○	○	○
Lack of understanding of media sector needs/requirements	○	○	○	○	○	○
Lack of communication & collaboration channels with the media sector	○	○	○	○	○	○
Compliance with relevant regulatory or legal frameworks	○	○	○	○	○	○
Lack of ethical guidelines/ framework for ethical conduct in own organisation	○	○	○	○	○	○
Lack of open access culture for developing and sharing AI tools	○	○	○	○	○	○
Lack of funding for the development of AI tools for media	○	○	○	○	○	○
Lack of promising AI talent that can work on these issues	○	○	○	○	○	○
US-China big tech monopoly in AI (facilitated by enormous funding & flexible regulations)	○	○	○	○	○	○

12. Are there other important challenges that AI researchers face when developing AI algorithms and tools for the media sector, which are not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Here you can elaborate on challenges that are specific to your own research area.

Enter your answer

Back | Next | Page 6 of 8

Adoption of AI solutions by media sector

In this section, we explore requirements for facilitating the adoption of AI technologies by the media sector.

13. What do you consider to be the easiest (or most promising) way for the adoption/development of AI tools for the media sector?
*

Only to be answered by those who understand these technologies and their potential. Please assess each option on a scale from 1 (difficult) to 5 (very easy).

	1	2	3	4	5	Don't know
Open source development tools	○	○	○	○	○	○
Ready-to-use components from open AI repositories	○	○	○	○	○	○
Enterprise software for dedicated AI tasks	○	○	○	○	○	○
Codevelopment with other media partners	○	○	○	○	○	○
Codevelopment with research partners	○	○	○	○	○	○
Cloud-based AI	○	○	○	○	○	○
Automated machine learning	○	○	○	○	○	○
Data science modeling tools	○	○	○	○	○	○
Code crowdsourcing	○	○	○	○	○	○

14. Are there other promising ways for the adoption/development of AI software/tools for the media sector that were not mentioned in the previous question?
This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

15. What kind of skilled personnel does a media company need to overcome the often encountered AI skills gap? *
You can select more than one option

- ☐ AI researchers
- ☐ AI software developers
- ☐ Data scientists
- ☐ Domain experts
- ☐ Business leaders able to interpret AI results
- ☐ Experts on AI ethics/legal issues
- ☐ Don't know
- ☐ Other

16. What are the most efficient ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry?
*

Please assess the efficiency of each option on a scale from 1 (not efficient) to 5 (very promising).

	1	2	3	4	5	Don't know
Sharing media industry data with the AI research community	○	○	○	○	○	○
Secondments/internships of AI researchers to the media industry	○	○	○	○	○	○
Research scholarships on media AI funded by the media industry	○	○	○	○	○	○
Collaboration of research labs and media industry on issues of common interest	○	○	○	○	○	○
Sub-contracting of AI media solutions to external research labs	○	○	○	○	○	○
Development of in-house AI research labs in media companies	○	○	○	○	○	○
Collaboration of media, academia and NGOs on topics with wide media or societal impact (e.g. AI for disinformation detection)	○	○	○	○	○	○
Organisation of challenges/benchmarks to provide novel AI solutions	○	○	○	○	○	○
Collaboration for the creation of open tools in open AI repositories	○	○	○	○	○	○
Training activities to increase AI skills in media industry	○	○	○	○	○	○

17. Are there other promising ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry that were not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back Next Page 7 of 8

AI ethics & AI regulation

In this section, we address ethical use of AI & AI regulations

18. On what matters would you like to have guidance from policymakers concerning use of AI systems in media? *

This is an open-ended question. Please elaborate.

Enter your answer

19. What aspects of AI use in media do you think should be regulated or further regulated? *

This is an open-ended question. Please elaborate.

Enter your answer

20. AI systems used in media present considerable opportunities and risks. For addressing these aspects, AI ethics is an important and useful tool to mitigate risks and increase benefits. What measures does your work environment have in place to control the ethical risks? *

- [ ] Ethical AI principles
- [ ] Ethical AI checklist
- [ ] Following or being a member of a Code on AI ethics
- [ ] Ethical board committee
- [ ] Ethics by design processes
- [ ] None
- [ ] I don't know
- [ ] Other

21. Have you ever done one of the following impact assessments?

Please select the assessments you have made.

- [ ] Data protection impact assessment (DPIA)
- [ ] Human rights impact assessment (HRIA)
- [ ] Assessment list for trustworthy AI (ALTAI)

Back | Submit | Page 8 of 8

15.2 AI4Media survey on AI for the Media Sector (version for media professionals)

AI4Media online survey on AI for the Media Sector

For media sector representatives as well as for those that their work is about the media sector more broadly (e.g. fact checkers that work in NGOs, representatives of independent authorities that supervise and regulate media, social scientists that study the media, etc.)

...

About this survey

About this survey
This questionnaire has been created in the context of the **AI4Media H2020 research project** (https://www.ai4media.eu/) and aims to help the AI4Media consortium in the development of a **Roadmap on Artificial Intelligence technologies and applications for the Media Sector**. Our aim is to collect the opinions of media sector representatives (news industry, film/TV/radio industry, gaming industry, etc) with regard to the most important trends, benefits, challenges, risks, and facilitators for the use of AI in the media sector. A similar questionnaire will also be shared with the AI research community. The results of this survey will be published as part of AI4Media's public deliverable "*D2.3 – AI technologies and applications in media: State of Play, Foresight, and Research Directions*".

The survey is **anonymous**, so there is no need to give your name or the name of your organisation. Data is not linked to respondents and is only going to be used in an aggregated way in this deliverable or any subsequent AI4Media publications.

Survey structure
The survey includes the following sections: (1) Professional backgound of respondents; (2) AI benefits for the media sector; (3) AI risks for the media sector; (4) AI challenges for the media sector; (5) AI strategies & skills in media sector; (6) Adoption of AI solutions by media sector; (7) AI ethics & regulation.

Each section usually includes 1-2 multiple choice questions + complementary open-ended questions (to elaborate, in case the previous choices are not enough). The questionnaire takes around **10-15 minutes** to complete.

Some survey questions and answers have been inspired by a) Deloitte's survey on "State of AI in the Enterprise", 2nd edition (https://tinyurl.com/2p8kzptv), and b) McKinsey's Global Survey on "The state of AI in 2020" (https://tinyurl.com/yckmce3n).

About AI4Media
Motivated by the challenges, risks and opportunities that the widespread use of AI has brought to the media, society and politics, the EU-funded AI4Media project aspires to establish a centre of excellence and a wide network of AI researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the Media sector, to make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and to reimagine AI as a beneficial enabling technology in the service of Society and Media. The project also includes a funding framework through open calls, an AI Doctoral Academy, and seven use cases to demonstrate the impact on the media sector.

More information about AI4Media can be found at https://www.ai4media.eu/. You can also watch our 3-min concept video on YouTube (https://tinyurl.com/2p9yrw3f).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911.

Personal data collection
No personal data of the survey respondents will be collected or stored. Answers to this survey will only be used in an aggregated way for the purposes described above.

Contact info
In case of questions about this survey, please send an email to filareti (at) iti (dot) gr

Consent
By clicking on **SUBMIT** after you have completed the last section of this questionnaire, you confirm that the information you have received has been sufficient and that you agree with the way this survey is being conducted.

Thank you for your participating in the survey! Your feedback is important.

Next

Page 1 of 8

Professional background

Professional backgound of survey respondents

1. What is your main professional background? *

- ◯ Journalist
- ◯ Newsroom staff member
- ◯ Fact checker or verification specialist
- ◯ Media regulator
- ◯ Music creator
- ◯ Film/TV industry creator
- ◯ Radio producer
- ◯ Podcaster
- ◯ Game developer
- ◯ Social media developer
- ◯ Content creator
- ◯ Content provider
- ◯ Advertising/marketing expert
- ◯ Sales manager
- ◯ CEO/business leader
- ◯ Innovation or R&D manager
- ◯ Media technology/Project manager
- ◯ AI researcher
- ◯ AI developer
- ◯ AI ethics/legal expert
- ◯ Social scientist
- ◯ Political scientist
- ◯ Other

2. What type of organisation do you work for? *

- ◯ Academic institution
- ◯ Research institution
- ◯ Public Service Media (PSM)
- ◯ Media & entertainment industry
- ◯ ICT industry
- ◯ Non-governmental organization (NGO)
- ◯ Independent authority
- ◯ Public sector
- ◯ Other

3. Please indicate the kind of media or media-related organisation you work for. *

- ◯ Media network (incl. TV, Radio, Online Media and News)
- ◯ TV station
- ◯ Radio station
- ◯ Streaming service
- ◯ Online news platform
- ◯ News/press agency
- ◯ Newspaper
- ◯ Magazine
- ◯ Fact-checking organisation
- ◯ Media regulatory authority
- ◯ Film industry
- ◯ Music industry
- ◯ Gaming industry
- ◯ Publishing industry
- ◯ Social media network
- ◯ Advertisement/marketing industry
- ◯ Content providers
- ◯ Other

4. In which country do you conduct your main professional or research activity? *

Please write the name of the country where your (main) affiliated organisation is located.

Enter your answer

Back | Next | Page 2 of 8

AI benefits for the media sector

In this section, we explore the benefits that AI technology can bring to the media sector.

5. What are the main benefits that AI can bring to the media sector in the next decade?
*

Only to be answered by those who understand these concepts/benefits. Please assess the importance of each benefit on a scale from 1 (not important) to 5 (very important).

	1	2	3	4	5	Don't know
Personalisation of content & services	○	○	○	○	○	○
Automation & optimisation of routine tasks	○	○	○	○	○	○
Enhance current services & products	○	○	○	○	○	○
Create or enable new services, products, content	○	○	○	○	○	○
Enhanced decision making	○	○	○	○	○	○
Increased productivity & operational efficiency	○	○	○	○	○	○
Increased creativity	○	○	○	○	○	○
Big data analysis to identify business opportunities, maximise revenue or conduct journalistic research	○	○	○	○	○	○
Facilitate expansion to new markets / target new audiences	○	○	○	○	○	○
Better monetisation of collected knowledge & data (e.g. video archives)	○	○	○	○	○	○

6. Are there other benefits that AI can bring to the media sector in the next decade that are not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Here you can elaborate on benefits that are specific to the media sub-sector you work in.

Enter your answer

7. In your opinion, does the use of AI have a clear value or benefit for your media organisation? *

○ Yes

○ No

○ It's not clear to me

○ I don't know

Back | Next | Page 3 of 8

* Required

AI risks for the media sector

In this section, we explore the potential risks of AI technologies for the media sector.

8. What are the potential AI risks of top concern to media companies? *

Only to be answered by those who understand these risks. Please assess the concern for each risk on a scale from 1 (no concern) to 5 (great concern).

	1	2	3	4	5	Don't know
Failure of AI in a critical mission	○	○	○	○	○	○
Making the wrong strategic decisions based on AI	○	○	○	○	○	○
Erosion of customer trust & brand damage due to AI failures	○	○	○	○	○	○
Legal responsibility for decisions made by AI	○	○	○	○	○	○
Ethical risks of AI	○	○	○	○	○	○
Risk of regulatory noncompliance	○	○	○	○	○	○
Cybersecurity vulnerabilities of AI	○	○	○	○	○	○
IPR issues with regard to new content creation	○	○	○	○	○	○
High expectations & low return on investment	○	○	○	○	○	○
Issues related to editorial/creative practices or journalistic values	○	○	○	○	○	○

9. What are you personally most worried about when it comes to the adoption of AI by the media sector? Both as a citizen as well as a representative of the broader media sector (or media researcher).
*

Please assess the following risks on a scale from 1 (not worried) to 5 (very worried).

	1	2	3	4	5	Don't know
Automation that shrinks human capital	○	○	○	○	○	○
Automation that limits human creativity	○	○	○	○	○	○
Unethical use of AI with destructive impact on individuals and society	○	○	○	○	○	○
Increased reliance on automated AI decisions without human in the loop	○	○	○	○	○	○
AI being a black box with little explanation on how decisions are made	○	○	○	○	○	○
AI models & software being susceptible to attacks	○	○	○	○	○	○
AI models being biased against groups of people	○	○	○	○	○	○
Inability to distinguish between created and real content	○	○	○	○	○	○
Exploiting user profiling on different platforms to create super profiles	○	○	○	○	○	○
Government not being able to effectively regulate AI	○	○	○	○	○	○
AI increasing the power of media tech giants, leading to monopolies	○	○	○	○	○	○
AI further amplifying echo chambers	○	○	○	○	○	○
Exploitation of media AI to hinder fundamental human rights, like freedom of expression	○	○	○	○	○	○

10. Are there other potential risks from the adoption of AI by the media sector, which are not mentioned in the previous questions?

This is an open-ended question aiming to complement the previous questions. Here you can also elaborate on risks that are specific to the media sub-sector you work in.

Enter your answer

Back Next Page 4 of 8

AI challenges for the media sector

In this section, we explore the challenges of developing AI technology for the media sector.

11. What are the most important challenges that the media sector faces when it comes to the adoption of AI?
*

Please assess each challenge on a scale from 1 (not important) to 5 (very important).

	1	2	3	4	5	Don't know
Understanding what AI can offer to a media company	○	○	○	○	○	○
Integrating AI into the company's work processes	○	○	○	○	○	○
Proving/measuring business value	○	○	○	○	○	○
Lack of relevant skills in personnel	○	○	○	○	○	○
Difficulties in attracting AI talent	○	○	○	○	○	○
Cost of developing AI solutions	○	○	○	○	○	○
Implementation challenges for AI solutions	○	○	○	○	○	○
Data privacy, access and integration	○	○	○	○	○	○
Understanding how AI tools work/make decisions	○	○	○	○	○	○
Ensuring compliance with relevant regulatory frameworks	○	○	○	○	○	○
Establishing an internal framework for ethical use of AI	○	○	○	○	○	○
US-China big tech monopoly in AI (facilitated by enormous funding & flexible regulations)	○	○	○	○	○	○

12. Are there other important challenges that the media sector faces when it comes to the adoption of AI, which are not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Here you can elaborate on challenges that are specific to the media sub-sector you work in.

Enter your answer

Back | Next | Page 5 of 8

AI strategies and AI skills in media sector

In this section, we explore what kind of AI strategies media companies adopt

13. Has your organisation invested in AI during the last year? *

- ◯ Yes
- ◯ No
- ◯ I don't know
- ◯ Other

14. Does your organisation have a specific AI strategy in place? If yes, what elements does your organisation's AI strategy include? *

If your organisation has no AI strategy in place, please only check the first option below and leave the rest unchecked.

- ☐ No specific AI strategy in place
- ☐ Clearly defined AI vision and strategy
- ☐ Roadmap prioritizing AI initiatives linked to business value
- ☐ Set of KPIs to measure the impact of AI initiatives
- ☐ Program to develop AI partnerships (e.g. with ICT companies, academia)
- ☐ Program for the recruitment of AI talent
- ☐ Standard tools and development processes in place for developing AI models
- ☐ Clear data strategy that supports and enables AI
- ☐ Scalable internal processes for labeling AI training data
- ☐ Clear framework to manage AI-related risks
- ☐ High-performance computing cluster for AI workloads
- ☐ Other

15. Should media companies train their own personnel to acquire AI skills or recruit new personnel with such skills (AI researchers, data scientists, etc.)? *

- ◯ Train media personnel to acquire AI skills
- ◯ Recruit experts with AI/data skills
- ◯ A new profile of media jobs will eventually emerge, combining both media and AI skills
- ◯ I don't know
- ◯ Other

16. What kind of skilled personnel does a media company need to overcome the often encountered AI skills gap? *

You can select more than one option

- ☐ AI researchers
- ☐ AI software developers
- ☐ Data scientists
- ☐ Domain experts
- ☐ Business leaders able to interpret AI results
- ☐ Experts on AI ethics/legal issues
- ☐ I don't know
- ☐ Other

Back | Next | Page 6 of 8

Adoption of AI solutions by media sector

In this section, we explore the adoption of AI technologies by the media sector.

17. What kind of ML/AI technologies do you mostly use in your organisation? *

- ☐ Image/video analysis
- ☐ Audio analysis
- ☐ Natural language processing
- ☐ Social network analysis
- ☐ User experience / audience analysis
- ☐ User profiling & recommendation systems
- ☐ Market analysis & forecasting
- ☐ Automatic content creation
- ☐ Content moderation
- ☐ Fact-checking & verification tools
- ☐ Automatic decision-making for business
- ☐ Other

18. What do you consider to be the easiest (or most promising) way for the adoption/development of AI tools for the media sector? *

Only to be answered by those who understand these technologies and their potential. Please assess each option on a scale from 1 (difficult) to 5 (very easy).

	1	2	3	4	5	Don't know
Open source development tools	○	○	○	○	○	○
Ready-to-use components from open AI repositories	○	○	○	○	○	○
Enterprise software for dedicated AI tasks	○	○	○	○	○	○
Codevelopment with other media partners	○	○	○	○	○	○
Codevelopment with research partners	○	○	○	○	○	○
Cloud-based AI	○	○	○	○	○	○
Automated machine learning	○	○	○	○	○	○
Data science modeling tools	○	○	○	○	○	○
Code crowdsourcing	○	○	○	○	○	○

19. Are there other promising ways for the adoption/development of AI software/tools for the media sector that were not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

20. What are the most efficient ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry?
*

Please assess the efficiency of each option on a scale from 1 (not efficient) to 5 (very promising).

	1	2	3	4	5	Don't know
Sharing media industry data with the AI research community	○	○	○	○	○	○
Secondments/internships of AI researchers to the media industry	○	○	○	○	○	○
Research scholarships on media AI funded by the media industry	○	○	○	○	○	○
Collaboration of research labs and media industry on issues of common interest	○	○	○	○	○	○
Sub-contracting of AI media solutions to external research labs	○	○	○	○	○	○
Development of in-house AI research labs in media companies	○	○	○	○	○	○
Collaboration of media, academia and NGOs on topics with wide media or societal impact (e.g. AI for disinformation detection)	○	○	○	○	○	○
Organisation of challenges/benchmarks to provide novel AI solutions	○	○	○	○	○	○
Collaboration for the creation of open tools in open AI repositories	○	○	○	○	○	○
Training activities to increase AI skills in media industry	○	○	○	○	○	○

21. Are there other promising ways to strengthen the transfer of AI knowledge, tools and best practices from academia/research to the media industry that were not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back | Next | Page 7 of 8

AI ethics & AI regulation

In this section, we address ethical use of AI & AI regulations

22. On what matters would you like to have guidance from policymakers concerning use of AI systems in media? *

This is an open-ended question. Please elaborate.

Enter your answer

23. What aspects of AI use in media do you think should be regulated or further regulated? *

This is an open-ended question. Please elaborate.

Enter your answer

24. AI systems used in media present considerable opportunities and risks. For addressing these aspects, AI ethics is an important and useful tool to mitigate risks and increase benefits. What measures does your work environment have in place to control the ethical risks? *

- [ ] Ethical AI principles
- [ ] Ethical AI checklist
- [ ] Following or being a member of a Code on AI ethics
- [ ] Ethical board committee
- [ ] Ethics by design processes
- [ ] None
- [ ] I don't know
- [ ] Other

25. Have you ever done one of the following impact assessments?

Please select the assessments you have made.

- [ ] Data protection impact assessment (DPIA)
- [ ] Human rights impact assessment (HRIA)
- [ ] Assessment list for trustworthy AI (ALTAI)

Back | Submit | Page 8 of 8

15.3 AI4Media survey on Media AI in the service of Society & Democracy

AI4Media online survey on AI in the service of Society & Democracy

About this survey

About this survey

This questionnaire has been created in the context of the **AI4Media H2020 research project** (https://www.ai4media.eu/) and aims to help the AI4Media consortium in the development of a **Roadmap on Artificial Intelligence technologies and applications for the Media sector**. Our aim is to collect the opinions of the AI research community and representatives of the broader media industry with regard to the most important benefits and risks that the use of AI for the media brings to the society and democracy. The results of this survey will be published as part of AI4Media's public deliverable "*D2.3 – AI technologies and applications in media: State of Play, Foresight, and Research Directions*".

Survey structure

The survey includes the following sections: (1) Professional backgound of survey respondents; (2) AI benefits for society & democracy; (3) AI risks for society & democracy; (4) Policies for ethical use of media AI & safeguarding of human rights.

Sections 2 and 3 each include one multiple choice question + one complementary open-ended question (to elaborate, in case the previous choices were not enough) while section 4 includes three multiple choice questions + one open-ended question. The questionnaire takes **about 5 minutes** to complete.

Answers to question no 4 were inspired by Khari Johnson's article on "How AI can empower communities and strengthen democracy", published in VentureBeat (https://tinyurl.com/235tnuys).

About AI4Media

Motivated by the challenges, risks and opportunities that the widespread use of AI has brought to the media, society and politics, the EU-funded AI4Media project aspires to establish a centre of excellence and a wide network of AI researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the Media sector, to make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and to reimagine AI as a beneficial enabling technology in the service of Society and Media. The project also includes a funding framework through open calls, an AI Doctoral Academy, and seven use cases to demonstrate the impact on the media sector.

More information about AI4Media can be found at https://www.ai4media.eu/. You can also watch our 3-min concept video on YouTube (https://tinyurl.com/2p9yrw3f).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911.

Personal data collection

No personal data of the survey respondents will be collected or stored. Answers to this survey will only be used in an aggregated way for the purposes described above.

Contact info

In case of questions about this survey, please send an email to filareti (at) iti (dot) gr

Consent

By clicking on **SUBMIT** after you have completed the last section of this questionnaire, you confirm that the information you have received has been sufficient and that you agree with the way this survey is being conducted.

Thank you for your participating in the survey! Your feedback is important

Next

Page 1 of 5

* Required

Professional background

Professional backgound of survey respondents

1. What kind of work do you do? *

- ◯ Member of the AI research community
- ◯ Working in the extended media industry
- ◯ Other

2. What type of organisation do you work for?

- ◯ Academic institution
- ◯ Research institution
- ◯ Public Service Media (PSM)
- ◯ Media & entertainment industry
- ◯ ICT industry
- ◯ Non-governmental organization (NGO)
- ◯ Independent authority
- ◯ Public sector
- ◯ Other

3. Which country do you work in? *

Enter your answer

Back | Next | Page 2 of 5

Media AI benefits for society & democracy

In this section, we explore the potential of media-related AI technologies to benefit the society and democracy.

4. What are the most promising AI media-related advancements that have the potential to empower the citizenry and strengthen democracy? *

Only to be answered by those who understand these technologies and their impact/potential. Please assess each option on a scale from 1 (not promising) to 5 (very promising).

	1	2	3	4	5	Don't know
Open source intelligence to hold power acountable (e.g. open-source data mining tools used to crack the Panama papers)	○	○	○	○	○	○
AI for emancipation (e.g. the Masakhane NLP project, aiming to increase representation of African languages and culture in tech)	○	○	○	○	○	○
Bias detection in economic, social, legal etc. processes	○	○	○	○	○	○
AI for fact-checking and fighting disinformation	○	○	○	○	○	○
AI for facilitating public dialogue and democratic consensus (e.g. the pol.is tool for analyzing and understanding what communities think)	○	○	○	○	○	○
AI for accessibility (creating more equitable access for people with disabilities)	○	○	○	○	○	○
Crowd - AI collaboration for cultural heritage preservation	○	○	○	○	○	○

5. Are there other promising AI media-related advancements that have the potential to empower the citizenry and strengthen democracy that were not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back Next Page 3 of 5

Media AI risks for society & democracy

In this section, we explore the risks of media-related AI technologies for the society and democracy.

6. What are the most significant AI media-related risks that have the potential to be detrimental to society and democracy? *

Please assess each risk on a scale from 1 (low risk) to 5 (high concern).

	1	2	3	4	5	Don't know
AI bias against groups of citizens, e.g. women, minorities, low-income people, vulnerable groups, etc.	○	○	○	○	○	○
AI tools used to spread disinformation and fuel polarisation (e.g. bots, fake content generation)	○	○	○	○	○	○
Over-personalisation of media services leading to filter bubbles	○	○	○	○	○	○
Use of AI tools to limit fundamental rights like freedom of expression	○	○	○	○	○	○
Unauthorised profiling of citizens	○	○	○	○	○	○
Unexplainable AI making decisions without human in the loop	○	○	○	○	○	○
AI affecting employment patterns	○	○	○	○	○	○
AI-induced inequality (people with access to data vs. people with no access)	○	○	○	○	○	○

7. Are there other significant AI media-related risks that have the potential to be detrimental to society and democracy that were not mentioned in the previous question?

This is an open-ended question aiming to complement the previous question. Please elaborate, if necessary.

Enter your answer

Back Next Page 4 of 5

Policies for ethical use of media AI & safeguarding of human rights

8. In your opinion, what specific fundamental rights are not sufficiently safeguarded with regard to the use of AI in media applications? *

You can select more than one option.

- [ ] The right to privacy and private life
- [ ] The right to freedom of expression and information
- [ ] The right to freedom of assembly and of association
- [ ] The right to freedom of the arts and sciences
- [ ] The right to freedom of thought, conscience, and religion
- [ ] The right not to be discriminated
- [ ] The right to human dignity
- [ ] The right to access to a fair trial and effective remedy
- [ ] The right to property (including intellectual property)
- [ ] All of them
- [ ] I don't know
- [ ] Other

9. What kind of policies should States or international organisations put in place to ensure that AI media applications are respectful of fundamental rights? *

You can select more than one option.

- [ ] Regulations coming with fines for stakeholders developing or providing AI media applications (researchers, media companies, platforms)
- [ ] Independent authorities overseeing AI media applications' risks and opportunities and providing guidance
- [ ] Standards, code of practices and other self-regulation instruments
- [ ] Banning AI media applications when demonstrated to be considerably harmful to fundamental rights
- [ ] Training and education programs on AI media applications and ethics / fundamental rights respect
- [ ] Collaboration spaces for citizens, NGOs, media sector, ICT companies to discuss AI risks & mitigation measures
- [ ] I don't know
- [ ] Other

10. What kind of policies should media companies put in place to ensure that AI media applications are respectful of fundamental rights? *

You can select more than one option.

- [ ] Reveal information on their algorithms and data
- [ ] Publish yearly reports on how they use AI in media applications, including information about collected data/ethics handling/content recommendation/content moderation, etc.
- [ ] Empower users to have influence on the content recommended to them, data collected from them, etc.
- [ ] Ensure appeal rights and establish services for user questions and complaints, regarding the use of AI systems in media
- [ ] Commit to and implement codes of conduct, certification, labels for compliant and ethical use of AI media applications
- [ ] Enable independent research on their services to analyse potential impact and risks
- [ ] I don't know
- [ ] Other

11. Do you have any best practices or recommendations to share regarding AI media applications respectful of fundamental rights?

Enter your answer

Back | Submit | Page 5 of 5

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 951911

info@ai4media.eu www.ai4media.eu