AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

D2.6

Final policy recommendations and Strategic Research Agenda on AI for the media industry

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	Final policy recommendations and Strategic Research Agenda on AI for the media industry
Deliverable number	D2.6
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	31 August 2024
Actual date of delivery	30 August 2024
Deliverable filename	D2.6_Final policy recommendations and Strategic Research Agenda on AI for the media industry_FINAL
Nature of deliverable	Report
Dissemination level	Public
Number of pages	183
Work Package	2
Task(s)	T2.2, T2.3
Partner responsible	KUL
Author(s)	Lidia Dutkiewicz (KUL), Noémie Krack (KUL)*, Filareti Tsalakanidou (CERTH) *This research is conducted under the supervision of Prof. Peggy Valcke and Dr Aleksandra Kuczerawy.
Editor	Lidia Dutkiewicz (KUL)
EC Project Officer	Evangelia Markidou

Abstract	Deliverable D2.6 presents i) the updated (final) version of the AI4Media Strategic Research Agenda (SRA), and ii) the final policy recommendations on AI for media. The SRA presents the main research themes for AI in media, highlighting current challenges, the research directions that need to be pursued to address them, the relevant media industry applications, and the potential impact of this research. In addition, the updated SRA includes a new chapter dedicated to Generative AI and Large Language Models, that examines the legal, ethical, and societal implications of Gen AI, as well as technology trends and applications for the media industry. The deliverable also presents policy recommendations for AI in media, addressing: (1) The need to impose and support good practices for transparency; (2) The need to support research in and of AI

	solutions and to stimulate responsible development of AI; (3) The need to mitigate AI divides and power imbalances; (4) The need for global and societally focused policies.
Keywords	Artificial Intelligence (AI), media, ethical AI, ChatGPT, AI Act, GDPR, DSA, generative AI, recommender systems, policy recommendations, media law, Strategic Research Agenda (SRA), Large Language Models (LLMs)

Copyright

Contributors

NAME	ORGANISATION
Lidia Dutkiewicz	KUL
Noémie Krack	KUL
Anna Schjøtt Hansen	UvA
Rasa Bocyte	NISV
Filareti Tsalakanidou	CERTH
Yiannis Kompatsiaris	CERTH

Peer Reviews

NAME	ORGANISATION
Rémi Mignot	IRCAM
Jens Fisseler	FhG

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	12/04/2024	Lidia Dutkiewicz, Noémie Krack	Final table of contents
0.2	07/08/2024	Noémie Krack	Review of the first draft version
0.3	09/08/2024	Lidia Dutkiewicz	Draft version sent to reviewers
0.4	23/08/2024	Filareti Tsalakanidou, Rémi Mignot, Jens Fisseler	Internal review
0.5	28/08/2024	Lidia Dutkiewicz	Integration of reviewers' comments
1.0	30/08/2024	Lidia Dutkiewicz	Final version sent to project coordinator

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf.

Table of Abbreviations and Acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
API	Application Programming Interface
Art.	Article
D.	Deliverable
DSA	Digital Services Act
DMA	Digital Markets Act
EC	European Commission
EMFA	European Media Freedom Act
EU	European Union
GDPR	General Data Protection Regulation
GenAI	Generative AI
GPAI	General-purpose AI
GPT	Generative Pre-trained Transformer
IIC	International Institute of Communications
KPI	Key Performance Indicator
LLM	Large Language Model
LMM	Large Multimodal Model
ML	Machine Learning
OS	Open Source
SRA	Strategic Research Agenda
T.	Task
T&C	Terms and Conditions
VLM	Vision Language Model
VLOPs	Very Large Online Platforms
VLOSEs	Very Large Search Engines
WP	Work Package

Index of Contents

1 Executive Summary 9

2 Introduction 12

3 AI4Media Strategic Research Agenda on AI for the Media Industry - updated version 19

3.1 Short summary of SRA contents 19

3.2 New section focusing on GenAI and LLMs 21

3.3 Updated SRA booklet 23

4 Policy Recommendations for AI in the Media Sector 25

4.1 Engagement with policymakers 25

4.2 Road towards final Policy Recommendations for AI in the Media Sector 29

4.3 Final Policy Recommendations for AI in the Media Sector 31

4.3.1 Impose and support good practices for transparency 31

4.3.2 Support research in and of AI solutions and stimulate responsible development of AI 36

4.3.3 The need to mitigate AI divides and power imbalances 39

4.3.4 The need for global and societally focused policies 42

5 Conclusions 48

6 References 50

7 Appendix: Updated AI4Media Strategic Research Agenda on AI for media (booklet) 54

Index of Figures

Figure 1: Challenges for media companies and initial recommendations 14
Figure 2: Challenges for Academia and Researchers and initial recommendations 14
Figure 3: Legal and Societal challenges and initial recommendations 15
Figure 4: The booklet of the updated AI4Media Strategic Research Agenda. 24
Figure 5: Panel Emerging realities: AI's impact on truth, disinformation and our perception of reality 26
Figure 6: Session 8: AI use in media – recommendations for policymakers 27
Figure 7: Roundtable Experts' perspective on policy & regulation 29
Figure 8: Hoodline's "AI" label appearing next to fake author personas on the site has been criticized for being an empty gesture 34
Figure 9: Recommendations towards good practices for AI transparency 36
Figure 10: Recommendations for supporting research in and of AI solutions and stimulating responsible development of AI 39
Figure 11: Recommendations to mitigate AI divides and power imbalances 42
Figure 12: Recommendations to support global and societally focused policies 47

1 Executive Summary

Deliverable D2.6 presents i) the updated (final) version of the AI4Media Strategic Research Agenda (SRA), and ii) the final policy recommendations on AI for media.

More specifically, we first offer an overview of the **updated version of the AI4Media Strategic Research Agenda (SRA).** A first version of the SRA was released in March 2023 and was presented in D2.7. An updated version was released in July 2024, aiming to reflect the significant progress made in the last year and a half with regard to Generative AI (GenAI) and Large Language Models (LLMs). To this end, a new section has been added to the AI4Media SRA that aims to address current technical or application limitations of GenAI and LLMs for the media but also deal with critical ethical and societal concerns. The section explores both technical challenges like domain adaptation, multimodality, explainability, AI hallucinations, resource-efficient models, dataset biases and fairness but also ethical, legal and societal aspects like AI policy & regulation, copyright of AI inputs and outputs, power dynamics, labor displacement, impact on human creativity, disinformation, environmental impact, and more. At the same time, important applications of GenAI for the media industry are discussed like use of LLMs as personal assistants, interactive agents in virtual worlds, synthetic content detection, artistic creation, etc.

Second, the deliverable presents the set of **final recommendations for the use of AI in the media sector**, adding to the initial recommendations presented in D2.4. The final recommendations cover aspects such as AI transparency, responsible AI development, power imbalances between tech and media, etc. and are briefly summarized below.

Recommendations to impose and support good practices for transparency:

1. We encourage strategies to support the **closure of the intelligibility gap** in the media sector and argue for sustainable funding to support the AI literacy of media staff.
2. We encourage the AI Office to draw a **code of practice** to facilitate the effective implementation of the obligations regarding the detection and **labeling** of AI-generated or manipulated content under Article 50 (7) AI Act.
3. The AI Office (and the Member States) could also use its (their) mandate in Article 95 AI Act to facilitate the drawing up of codes of conduct concerning the voluntary application of specific requirements of the AI Act – such codes of conduct could relate to best **AI transparency** practices.
4. Article 96 (1)(d) AI Act obliges the European Commission to develop **guidelines** on the **practical implementation of the transparency** obligations laid down in Article 50 AI Act. When issuing such guidelines, the EC should pay particular attention to the needs of small and medium-sized enterprises including start-ups, of local public authorities, and of the sectors most likely to be affected by the obligations of Article 50 AI Act. We encourage the EC to be in **dialogue with the media stakeholders** to take into account the needs of the media sector.

5. In the lack of harmonized standards and common specifications on labeling, we encourage media organizations, associations, and relevant media authorities to work together towards a **common approach to labeling the use of AI in the news** (and media) sector.
6. We encourage policies that require increased **third-party providers of AI solutions transparency**. We opt for more support and guidance for media organizations to guide the procurement process and for scrutiny of whether the sellers of the AI solutions share the relevant information.

Recommendations to support research in and of AI solutions and stimulate responsible development of AI:

1. We encourage the EC to **clarify** in its Delegated Act on data access the exact **scope**, **conditions**, and **mechanisms** for **data access** for researchers, including also (investigative) journalists.
2. We support policies promoting **publicly available datasets** and open data access and open APIs for research or investigative purposes (e.g., training data or algorithmic accountability reporting) which go beyond DSA's 'systemic risk'.
3. We encourage relevant authorities to stay vigilant about '**open washing**' and to make sure that data access gatekeepers do not circumvent the applicable provisions and provide meaningful insights into the workings of the AI model.
4. We encourage project funders to support and **fund data spaces for the media industry**, the development of open datasets and AI solutions for low-resource languages.
5. We call on funders to focus on **long-term funding schemes** for the benefit of society that prioritize sustainability over short-term quick results.

Recommendations to mitigate AI divides and power imbalances:

1. We encourage national authorities to establish **regulatory sandboxes** to enable testing and experimentation for media organizations.
2. We support the initiative to create **LLMs in European regional and minority languages**.
3. We encourage the creation of standardized **frameworks for model openness** assessment made up of multiple elements and we encourage national authorities to invest in the funding schemes that can help enable regional and local collaborations on AI that can allow for the development of scalable and open-source AI solutions to be used in multiple organizations.
4. We encourage **alternative funding schemes** that allow organizations developing AI to be independent of funding from big tech companies.
5. We suggest strategies which **support media organizations in AI integration** and the development of localized solutions, independently from big tech providers.

Recommendations for globally and societally focused policies:

1. We encourage the development of **opt-out compliance policies** and the establishment of generally recognized standards or protocols for the machine-readable expression of the rights reservation and opt-outs from training generative AI models.
2. We encourage initiatives which focus on the **collective responses of media content creators to data scraping** for model training.
3. We encourage the AI Office to adopt a **standard template** on what the "sufficiently detailed summary" provided by GPAI developers means; such template should in particular contain elements regarding the ethical review process of the training data, its collection, legal basis, its diversity and whether the data was scraped from the internet and information about the crawling method.
4. We encourage the Member States to use their mandate in Article 95 AI Act to develop **codes of conduct for the assessment and minimizing of the environmental impact** of AI systems.

2 Introduction

This deliverable (D2.6) presents the updated versions of i) the AI4Media Strategic Research Agenda on AI for the Media Industry (D2.7) that was published in March 2023, and ii) the Pilot Policy Recommendations for the use of AI in the Media Sector (D2.4), which were published in August 2022.

AI4Media Strategic Research Agenda

AI4Media's vision is that of a European Network of Excellence in AI for the Media, Society and Democracy that will glue together the pieces of the currently fragmented European AI landscape and promote a unique brand of European Media AI. AI4Media builds a network of experts, including both leading researchers in media AI from academia and research as well as top European media companies that use AI to enhance their operations and business opportunities. Together, they collaborate to address significant technical, legal, ethical and application challenges, aiming to address pressing needs of the media industry and significant societal problems.

The delivery of the AI4Media Strategic Research Agenda is a key contribution towards building this European Network of Excellence focusing on AI for the Media and Society. Based on the work carried out for D2.3 "AI technologies and applications in media: State of Play, Foresight, and Research Directions", the first version of the SRA (D2.7) was delivered in March 2023 and laid out the strategic plan for AI4Media's R&I activities, presenting the main challenges to be tackled, the research directions that need to be pursued to address them, the relevant media industry applications, and the potential impact of this research.

The SRA comprised the following elements:

- An AI for Media Observatory to monitor the AI policy/regulatory landscape, AI societal impact, and AI technology trends, providing insights, analyses and recommendations for policymakers, industry and researchers.
- Research on four core AI areas that help reinforce and extend Europe's expertise in AI for Media, including new machine learning paradigms, trustworthy AI, content-centered AI, and human/society-centered AI.
- Real-world case studies showcasing how to transform AI research into practical applications for the media industry with concrete impact to society and the economy.
- AI education and AI skills development through the establishment of a prestigious European PhD programme on AI (i.e. the International AI Doctoral Academy - AIDA), a flexible mobility program (i.e. the Junior Fellows Exchange Program), and open calls for funding research labs and SMEs working on media AI (i.e. the AI4Media open calls).

Following the third review recommendations, an updated version of the SRA was delivered in July 2024, aiming to reflect on the recent rapid advancements in the field of Generative AI. This new wave of AI innovation offers amazing opportunities for the media but at the same time raises significant concerns about its impact on creative industries, society, democracy, environment, etc. These aspects are explored in a newly added section dedicated to GenAI and LLMs, which examines i) legal, ethical and societal aspects (e.g. AI policy & regulation, copyright of AI inputs and outputs, power dynamics, labor displacement, impact on human creativity, disinformation, environmental impact), ii) technology aspects (e.g. domain adaptation, multimodality, explainability, AI hallucinations, resource-efficient models, dataset biases and fairness), and iii) relevant media and creative industry applications (e.g. LLMs as personal assistants, interactive agents in virtual worlds, synthetic content detection, artistic creation, etc.).

The SRA aims to become a useful tool and point of reference for researchers, media practitioners and policymakers working on the intersection of AI and media. The relevant booklet is available on the project website[1].

Policy Recommendations for the use of AI in the Media Sector

The pilot policy recommendations of D2.4 represented a first attempt to provide guidance on how to better regulate AI in the context of the media sector. They were based on previous deliverables, including the initial whitepaper on the social, economic, and political impact of media AI technologies (D2.2), which also identified specific policy recommendations, as well as a public survey on AI for the media addressed to AI researchers and media industry professionals (D2.3) and a public stakeholders' consultation targeting the research community, media industry and policymakers.

As a reminder, below we provide the overview of the most recurring challenges and corresponding recommendations identified in D2.4 (Figure 1, Figure 2, and Figure 3).

[1] AI4Media Strategic Research Agenda (latest version): https://www.ai4media.eu/strategic-research-agenda-on-ai-for-the-media-industry/https://www.ai4media.eu/strategic-research-agenda-on-ai-for-the-media-industry/

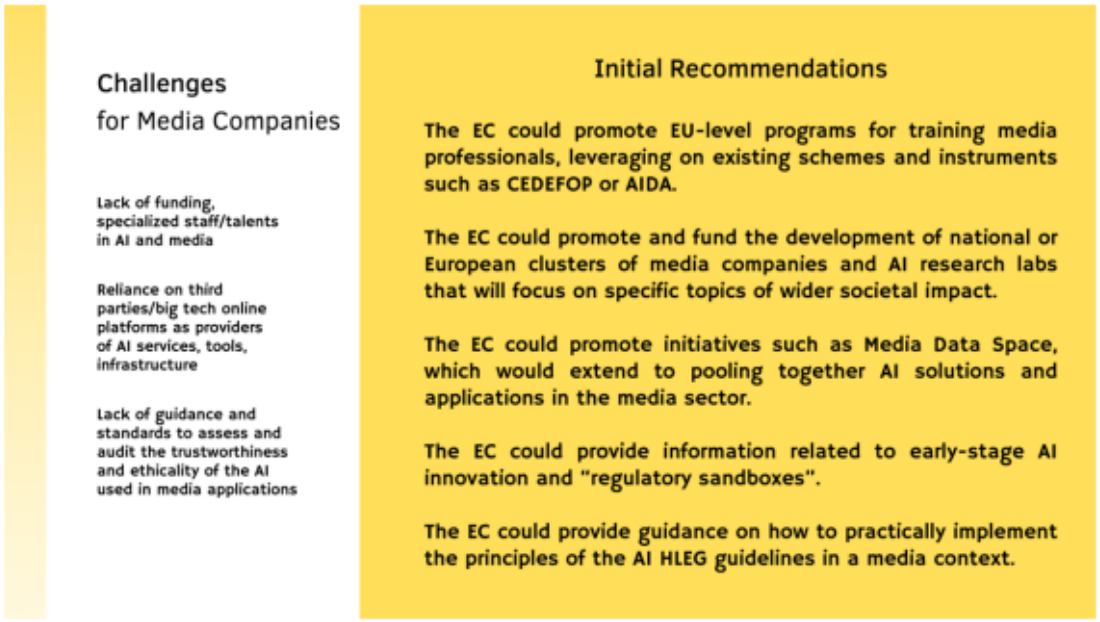


Figure 1: Challenges for media companies and initial recommendations

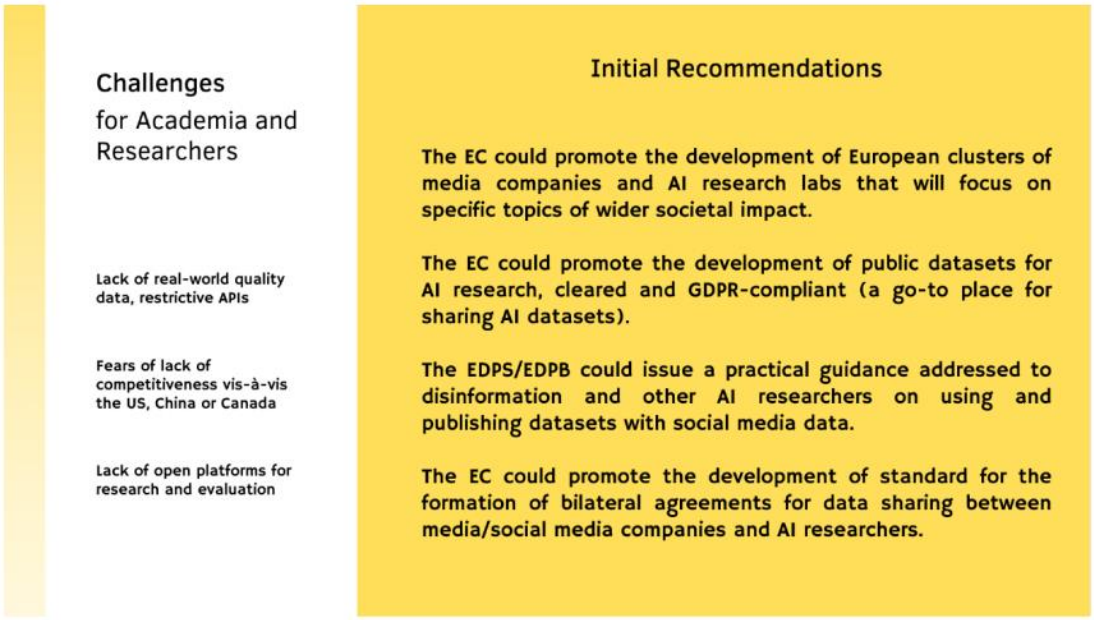


Figure 2: Challenges for Academia and Researchers and initial recommendations

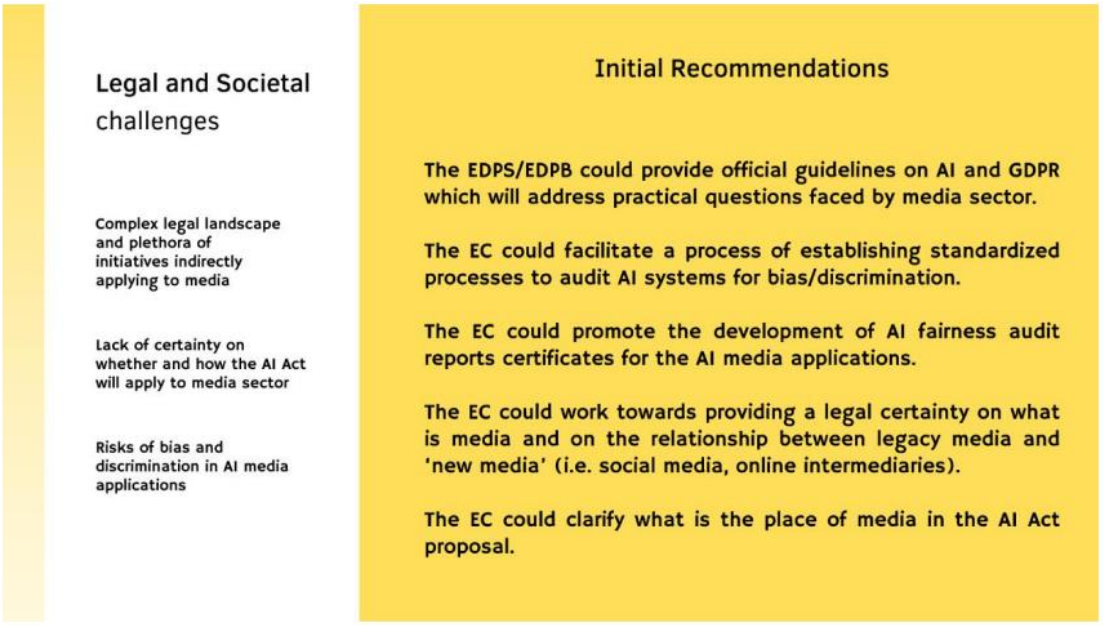


Figure 3: Legal and Societal challenges and initial recommendations

Between 2022 and 2023, three different **industry workshops** were conducted in the context of WP2 with three different stakeholder groups. The aim of these workshops was to disseminate and qualify the Pilot Policy Recommendations for the use of AI in the Media Sector. The workshops concerned the topics of (1) Measuring the success of recommender systems for media, (2) Identifying common challenges relating to the use of AI in content moderation, and (3) The current challenges and future paths of AI in audiovisual archives.

The methodology for each workshop was developed in close collaboration with partners of the consortium with expertise within the concrete challenge to ensure a relevant setup for each workshop. This included drawing particularly on the expertise of KUL, UvA and NISV, as well as expertise within WP2. The two online workshops focusing on recommender systems and content moderation were both two hours long and were invite-only, as the focus was on getting the right people in the room, which included media organizations, platforms, and AI developers. They also shared a similar setup, where the workshop started with an inspirational talk by a key expert on the topic, followed by a round table with the participants before proceeding with a discussion of potential good practices. The physical workshop focusing on AI in audiovisual archives took place over two days at the headquarters of HRT (Croatian Radiotelevision) in Zagreb, Croatia and included participants from seven organizations from the Balkan and Mediterranean regions. The decision to conduct an extended, in-person workshop in this region stemmed from a notable gap highlighted in the initial whitepaper concerning a general lack of knowledge about AI in audiovisual archives, and a substantial geographical disparity in the current literature on AI in the media sector.

For more details about the methodology, process and the outcomes see the Final white paper on the social, economic, and political impact of media AI technologies (D2.5).

Moreover, three other **policy workshops** were designed to engage the participants in active discussion over the legal challenges that were highlighted in the initial pilot recommendations and to offer them the chance to provide input on what policies they found to be missing from their vantage point in the sector. Oppositely to the industry workshops above, these policy workshops all followed the same format and instead focused on conducting the workshop with different stakeholders. The format was designed based on the structure of the pilot policy recommendations, which highlighted both challenges and recommendations for three areas, namely (1) Media companies, (2) Academia and researchers and (3) Legal and Societal.

The first policy workshop was conducted during the Joint Computation + Journalism European Data & Computational Journalism Conference 2023, in Zurich, Switzerland in June 2023, where the participants were representatives from media organizations or media researchers. The second workshop was conducted during the 9th General Assembly of the AI4Media consortium, in Pisa, Italy, in October 2023, where the participants consisted of consortium partners and, as a result, included many technical experts as well as use case partners. The third workshop was conducted with the AI, Media, and Democracy Lab at the University of Amsterdam, Netherlands in November 2023, where the participants included media and legal scholars working in the intersection of AI and media.

The workshops were divided into three parts. First, the participants were given a presentation of the Pilot Policy Recommendations for the use of AI in the Media Sector (D2.4). Second, they engaged in the first exercise titled '**Provotyping Policies'**. Participants were presented with three provocative statements that were aimed at highlighting specific tensions in the current AI and media landscape.

For the media landscape, the following statements were used:

- 'Generative AI should increasingly be used to write news'
- 'Editorial content from media organisations should never be removed by private platforms if not illegal'
- 'Media organisations should simply use easily accessible AI solutions (like OpenAI) rather than open source in-house development'

For the research landscape, the participants were asked to react to the following statements:

- 'API privatisation like on Twitter (now X) will become standard practice and research access will be lost'
- 'Training data becomes a problematic trade-off - either you cannot get the data you need due to strict regulation or you have to use illegal/unethical datasets'
- 'AI development/research should never rely on funding from big tech (e.g. Google DNI, Meta)'

For the regulatory and policy landscape, the following statements were suggested:

- 'Big tech wants strict regulation of AI to eliminate small competitors, such as small media organisations or start-ups'
- 'The use of AI should always be transparently disclosed - this should be a strict regulatory demand'
- 'All applications of AI in media should be high risk in the AI Act'

Third, they were introduced to a **'Wishing Well'** exercise where they were asked to first write down three wishes for policies that they felt were highly needed. As the last part of the exercise, they were asked as a group to rank each of the overarching themes from most important to least important and finally present this to the other groups.

Regulators have for years attempted to grasp and address the emergence of AI and the EU has taken regulatory steps to address the challenges and opportunities presented by AI. At the time the workshops were held, the AI Act had been discussed and amended. On 13 July 2024, the final version of the AI Act was published. Therefore, the final policy recommendations have a double purpose. First, we dive into these new regulations to explore how they address the media sector - or in some cases leave things unclear – and offer recommendations to remedy this.

We explore how particularly the AI Act and the Digital Services Act (DSA) are addressing six cross-cutting policy needs that are deemed central for supporting a responsible AI approach in the media sector. These six cross-cutting policy needs are a result of both desk research and several workshops conducted with media stakeholders, legal and media researchers as well as AI developers and researchers throughout the AI4Media project. They are described in detail in the '*Final white paper on the social, economic, and political impact of media AI technologies*' (D2.5) published in December 2023.

Second, the aim of the recommendations is also towards helping media organisations understand what legislation could support them. We wish to provide industry professionals with insights and findings they can leverage in policy discussions to point to weaknesses in the legislation that affect the media sector's ability to act responsibly in the AI landscape.

As the enforcement of the AI Act and the DSA has just began, we aim for the policy recommendations to provide interpretive guidance, suggestions and points of attention to, among others, the enforcers, national authorities and media councils.

Deliverable Structure

In **Section 3**, we report on the updated **AI4Media Strategic Research Agenda** released in July 2024, focusing on the new section discussing GenAI and LLMs. The updated SRA booklet is included in the Appendix (Section 7).

Then, **Section 4** of this deliverable offers the **final policy recommendations on AI for the media industry**. Section 4.1 illustrates the efforts to engage with policymakers to get insights into the final policy recommendations. Section 4.2 explains the road towards the final policy

recommendations on AI for media. Building on the series of workshops, desk research and the *Final white paper on the social, economic, and political impact of media AI technologies* (D2.5), we take as a starting point the following cross-cutting policy needs, which were deemed therein to be the most urgent and important based on their impact on the media industry: (1) The need to impose and support good practices for transparency; (2) The need to support research in and of AI solutions and to stimulate responsible development of AI; (3) The need to mitigate AI divides and power imbalances; (4) The need for global and societally focused policies. In Section 4, we therefore assess whether the current EU legal framework (the Digital Services Act, the AI Act, …) addresses these needs, and we propose policy recommendations where needed (Section 4.3).

Finally, **Section 5** offers conclusions.

3 AI4Media Strategic Research Agenda on AI for the Media Industry - updated version

A first version of the AI4Media Strategic Research Agenda (SRA) was released in March 2023 and was presented in D2.7, submitted in M36. In the current deliverable (D2.6), we present an updated version of the AI4Media SRA which was released in July 2024, aiming to reflect the significant progress made in the last year and a half with regard to Generative AI, fuelled by the widespread adoption of Large Language Models (LLMs) and the rise of Large Multimodal Models (LMMs). This new wave of AI innovation offers amazing opportunities for the media but at the same time raises significant concerns about its impact on creative industries, society, democracy, environment, etc. These aspects are explored in a new SRA section (section 8: Generative AI and LLMs for the Media) dedicated to Generative AI and LLMs that examines legal, ethical and societal aspects, key technology trends, and relevant media and creative industry applications.

3.1 Short summary of SRA contents

The Media are already benefiting from AI advancements and AI-driven applications that can significantly facilitate, enhance or transform important tasks, including smart assistants, smart recommender systems, content personalization, automatic content creation, multi-modal content search, multilingual translation, disinformation and manipulated content detection, social media analysis and trend detection, online debate analysis, forecasting and decision support-systems, and many more. Further advances in AI have the potential to transform the media industry and revolutionize how operations run and how content is created, delivered and consumed while they can also offer trustworthy solutions with a societal impact, aiming to improve political participation, increase social cohesion, equip citizens against disinformation, and encourage healthy debates and social interaction.

To realize this enormous potential of AI for the media industry requires breakthroughs in several domains such as:

- Machine learning (ML), aiming to address important challenges of current ML techniques, including learning with few data, learning on-the-fly, transfer of knowledge and optimal AI architectures. In addition, research should also focus on distributed AI systems running on heterogeneous devices but also disruptive technologies currently at the laboratory stage such as Quantum-assisted Reinforcement Learning.
- Content-centred AI technology, valuable for the media industry and marketable as end-user services, such as multimedia metadata extraction, summarization, and clustering, automatic audiovisual content generation and enhancement, linguistic analysis, and media-specific core technologies to improve learning performance.
- Human and society-centred AI technology, to equip citizens and media professionals with a set of tools that can be used to counter the effects of media manipulation and

disinformation, enhance the understanding of online debates, support the analysis of perceptions of social media and the effects of online data sharing, and improve local news understanding without being limited by language barriers.

- Trustworthy AI techniques, that aim at providing a framework for the development of the technologies mentioned above that guarantees their suitability with respect to democratic and ethical values. Research should focus on issues of robustness against threats and malicious attacks, explainability of AI decisions, fairness and mitigation of bias of AI models, and techniques for privacy-preserving AI.

In addition to the aforementioned topics of interest, another research area where rapid progress has been observed in the last year and a half is that of Generative AI. The wide adoption of Large Language Models like ChatGPT and the rise of large multimodal models is transforming existing media industry workflows, essentially reimagining content creation and consumption. Research should focus on both technical challenges like domain adaptation, multimodality, explainability, AI hallucinations, resource-efficient models, dataset biases and fairness but also on ethical, legal and societal aspects like AI policy & regulation, copyright of AI inputs and outputs, power dynamics, labour displacement, impact on human creativity, disinformation, environmental impact, and more. At the same time, important applications of generative AI for the media industry should be studied like use of LLMs as personal assistants, interactive agents in virtual worlds, synthetic content detection, artistic creation, etc.

These research outcomes are integrated and evaluated in real-world use cases, aiming to address significant challenges currently faced by different media industry sectors and to highlight how AI applies throughout the media industry value chain, from research and content creation to production, distribution, consumption/interaction, performance and quality measurement. The use cases cover a variety of media and societal topics such as disinformation, news research and production, organization of media archives and content moderation, game design, human-machine artistic co-creation, and social science research.

In parallel to delivering the next generation of AI research at the service of media, AI4Media established an AI Media Observatory to monitor the legal and technological landscape as well as the impact of media AI on the society, economy and democracy. The Observatory provides an overview of the existing EU policy and legal initiatives and their impact on future AI research for the media industry, analyses ethical, societal, environmental and economic concerns, and provides easy access to leading experts in this domain.

Implementing our vision of AI as a human-centred, trusted, and beneficial enabling technology in the service of media and society, requires supporting in practice the next generation of AI talent in Europe by offering opportunities for top AI education and skill development while also supporting entrepreneurship and innovative ideas. To this end, AI4Media established the International AI Doctoral Academy, a joint ICT-48 instrument to support world-level AI education and training for PhD/postdoc AI researchers. In addition, it provides mobility opportunities for

young researchers and media professionals. And lastly, it funds and supports SMEs, start-ups and research labs that want to develop innovative applications and research for the Media. These activities further strengthen the European AI research community.

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 and transform the industry. The AI4Media Network of Excellence aims to play an important role in this transformation, by bringing together leading research and industry players in this domain to strengthen the competitiveness and growth of the European media industry and increase Europe's innovation capacity in media AI. The updated version of this Strategic Research Agenda crystallises AI4Media's research and innovation activities to materialise this vision.

3.2 New section focusing on GenAI and LLMs

Content creation is the pillar of the media and entertainment industry, with media companies, broadcasters, journalists, artists, influencers, etc. making a living out of producing a large stream of multimedia data. While most of the professional media content is still manually edited and originated, the rise of Generative AI, mainly driven by the recent explosion of Large Language Models, is already drastically transforming the content creation process for both media companies and independent creators by automating tasks that require creativity.

Trained with massive amounts of data, Generative AI models are able to automatically produce new, original high-quality content in the form of text, image, audio, video, 3D representations, etc. that looks like the data the model was trained on. Large language models like GPT are essentially next-word prediction engines that process natural language inputs, known as prompts, and produce realistic text. Text-to-image models like DALL·E 3 can generate images that accurately match complex and nuanced text prompts. Vision language models (VLMs) like CLIP can learn simultaneously from images and text and are used in tasks like visual question answering, image search or image captioning.

The launch of ChatGPT in 2022 was a turning point that highlighted how GenAI can be mainstreamed to help both media companies and consumers to create new content. Since then, the wide adoption of LLMs and the recent rise of large multimodal models that can process and understand multiple types of data modalities, are creating amazing opportunities for the media industry. The range of potential applications is limitless: deepfakes for film/TV, film preservation, automatic news article production, music composition, realistic game asset development, automatic script generation, personalised advertisements, interactive storytelling, interactive virtual worlds etc.

At the same time, Generative AI presents significant challenges and risks for media companies, creators and consumers but also for the general public. Dataset biases, soaring environmental costs, lack of explainability for model outputs, intellectual property violation concerns,

deepfakes and disinformation, AI hallucinations, increasing power imbalance between large media companies / tech providers and smaller companies / independent AI developers and content creators, broader societal implications etc. are only some of the concerns that have emerged recently.

The new section of the SRA (section 8), summarises the main areas of research interest under the Generative AI and LLMs research theme. The aim is to efficiently address current technical or application limitations of GenAI for the media but also deal with critical ethical and societal concerns. We explore i) legal, ethical and societal aspects, ii) key technology aspects, and iii) media and creative industry applications of GenAI.

More specifically, the following topics are covered (see list below). For each topic, we present the relevant challenges, the research directions that should be pursued to tackle these challenges, and the potential impact of this research on the media, society and democracy.

- Research on legal, ethical and societal aspects
    - Generative AI and the AI Act
    - Dataset diversity and quality
    - Challenging power dynamics and infrastructure capture
    - AI-generated content and copyright laws
    - Impact on epistemic welfare and the (dis)information ecosystem
    - Impact of Generative AI in media production and creativity
    - New displacement patterns and hidden labour
    - Good practices for the use of GenAI in the media sector
    - Generative AI Literacy for media professionals and media consumers
    - Environmental impact
- Research on key technology aspects
    - Domain adaptation and fine-tuning of Large Models
    - Incorporating knowledge in foundation models
    - LLM hallucinations and LLM output quality
    - LLM unlearning
    - Embodied LLMs
    - Dataset biases and fairness
    - Enhancing the interpretability of Generative AI models
    - Explainable vision-language models
    - Resource-efficient vision-language models
    - Physics in vision-language models
    - Integration of new/multiple modalities in next-generation Large Models
    - Collection-wise Visual Question Answering
    - Generalizable evaluators for high-quality content generation
    - Face dataset anonymization
    - Enterprise context-aware Generative AI

- Media and creative applications
    - Personalised LLMs for journalists
    - LLMs as an assistant research tool for media analysis
    - Indexing the entire World Wide Web with LLMs embeddings
    - Media agents
    - Audio-visual synthetic media detection
    - Content provenance markers
    - Ensuring transparency and the right to be informed through labelling the use of Generative AI in media production
    - Real-time music generation
    - Emotionally expressive avatars
    - Multimodal text instructable agents in interactive worlds.

3.3 Updated SRA booklet

To facilitate the widest possible dissemination of the SRA, a booklet was designed for the first version of the SRA by LOBA and CERTH and was released in March 2023 (see the Appendix of D2.7). An updated version of the booklet was developed during this period to reflect the updated SRA content (see Figure 4). The updated SRA booklet was delivered in July 2024 and is available on the project website[2] as well as on the VISION CSA website[3].

The AI4Media SRA booklet is also included as an Appendix in this report (see Section 7).

The updated SRA booklet has been widely disseminated through all AI4Media communication channels as well as via ICT-48 and EC channels. In addition, a relevant blogpost was published in the European AI Alliance blog.[4]

Moreover, the contents of the updated AI4Media SRA were used to inform and update the joint ICT-48+2 Strategic Research Agenda on AI, data, and robotics 'made in Europe'.[5] This updated version of the joint SRA discusses how Generative AI impinges on the research challenges identified in the previous version of the joint SRA.

[2] AI4Media Strategic Research Agenda on AI for the Media industry: https://www.ai4media.eu/strategic-research-agenda-on-ai-for-the-media-industry/
[3] AI4Media SRA on VISION CSA website: https://www.vision4ai.eu/sra/
[4] Blogpost about the updated AI4Media SRA: https://futurium.ec.europa.eu/en/european-ai-alliance/blog/updated-ai4media-strategic-research-agenda-ai-media-industry-new-section-about-generative-ai
[5] Joint ICT-48+2 Strategic Research Agenda on AI, data, and robotics 'made in Europe': https://www.vision4ai.eu/sra/

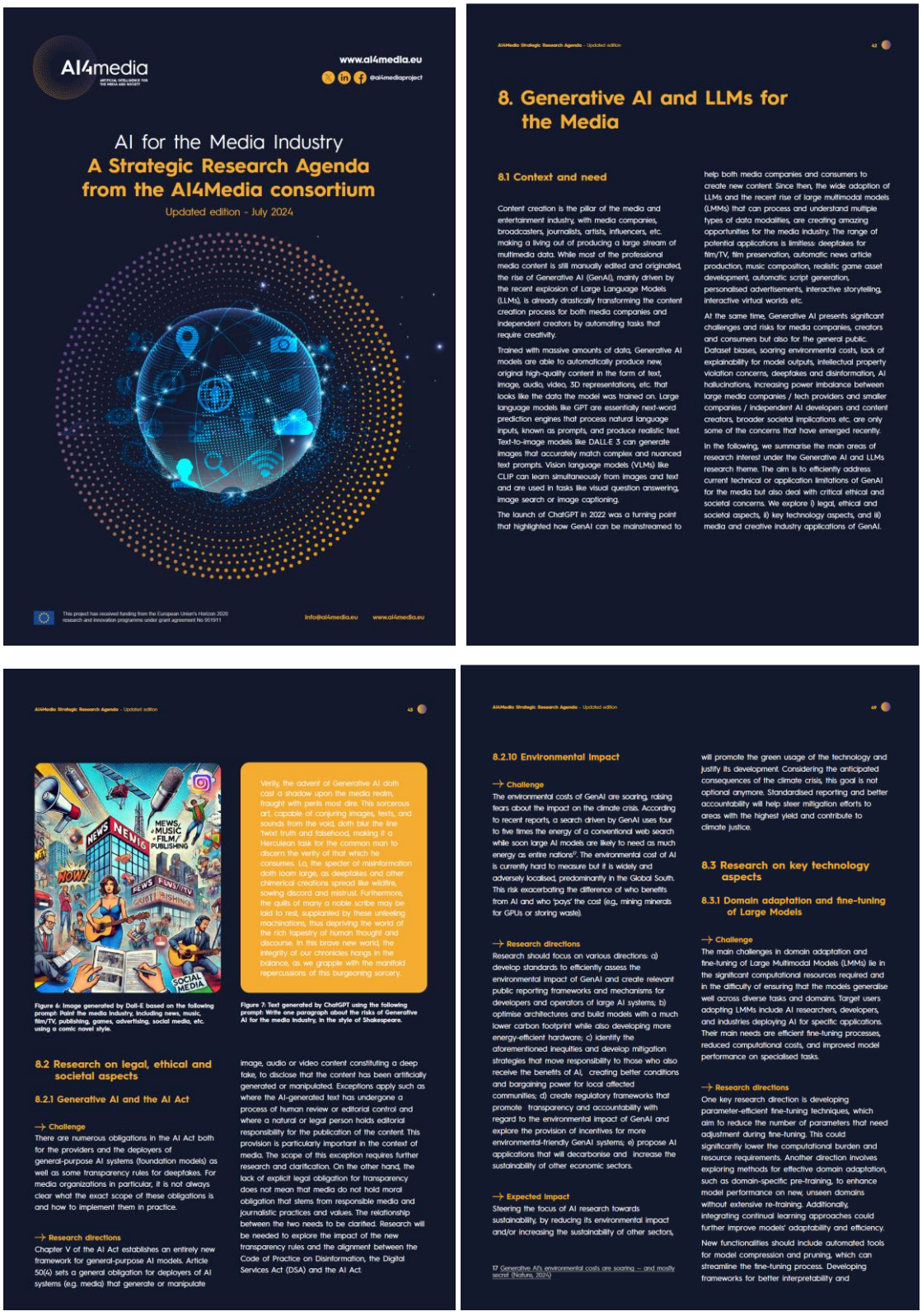


Figure 4: The booklet of the updated AI4Media Strategic Research Agenda.

4 Policy Recommendations for AI in the Media Sector

4.1 Engagement with policymakers

This section provides information on some of the main outreach activities undertaken to engage policymakers in the discussion about the final scope of the policy recommendations. These activities enabled us collect input to fine tune our analysis and work.

■ Meet the Future of AI events

AI4Media collaborated with the AI/Disinformation projects cluster on different occasions. The cluster includes the Horizon Europe projects vera.ai, AI4Trust, TITAN, AI4Debunk, and AI-CODE.

In June 2023, the projects organised an event in Brussels entitled "Meet the Future of AI: Countering Sophisticated & Advanced Disinformation". The event focused on relevant issues and challenges around generative AI and tackling disinformation. The event was an opportunity to raise awareness about the project's scope and focus on legal and policy aspects. A report about the event is available online.

A second edition of the 'Meet the Future of AI' event was organised in June 2024 in Brussels. The event was made possible thanks to the help of KU Leuven through their contacts with the Brussels Media Authorities who hosted the event. The event focused on Generative AI, disinformation, and democracy in this crucial year for elections around the globe. It shared some fresh research updates following the EU elections held in May 2024. A report about the event is available online.

In addition to these events focusing on legal and policy considerations of the use of AI in the media sectors, the projects in collaboration with EDMO joined forces and published a new white paper on Generative AI and Disinformation: Recent Advances, Challenges, and Opportunities. One of the focuses of the paper was to discuss the ethical and legal challenges associated with GenAI and disinformation. KUL contributed to the legal and policy aspects of the paper.

■ EU DisinfoLab annual conference

In October 2023, KUL participated in the annual conference of the EU DisinfoLab in Krakow. The event brought together nearly 400 renowned professionals from diverse backgrounds for a two-day event with different sessions formats. KUL provided an expert talk on the panel discussion: "Emerging realities: AI's impact on truth, disinformation and our perception of reality". The talk was the opportunity to present AI4Media and share some takeaways from the AI EU legal and policy analysis conducted during the project and inform about the forthcoming final policy recommendations. On this occasion, some AI4Media leaflets on WP2's work were distributed. A briefing about the importance of following-up with the implementation of EU legislation through the technical standards process was underlined. A call for the presence of civil society

and engagement from this counterpart to balance the interest of stakeholders involved in standardisation processes.

Figure 5: Panel Emerging realities: AI's impact on truth, disinformation and our perception of reality

- European Social Dialogue in the audio-visual sector

In November 2023, KUL provided a talk during a meeting of the European social Dialogue in the audio-visual sector. The EU's audiovisual sector is composed of public and commercial broadcasting, and independent television and film production. The dialogue regroups representatives from the workers and employers from the EU audiovisual sector including: the European Broadcasting Union, the European Federation of Journalists, the URO-MEI - UNI-Europa Media - Entertainment & Arts, and so forth.

The talk focused on Generative AI and the law but also presented the AI4Media, the policy recommendations and the AI Media Observatory. The Q&A and follow-up exchanges provided the opportunity to learn from the different representatives' concerns for the sector. Discussions among participants covered various issues such as:

- consent for data use (opt-out vs. opt-in)
- the harm of AI for the cultural sector
- the impact on employment in the audiovisual sector
- the need to address AI as a transversal issue and not in different silos, the media sector as critical infrastructure
- the need to ensure that algorithms are human-centred.

Minutes from the meeting were taken and sent internally to all members of the Dialogue afterwards, which enabled reaching the representatives which could not attend the physical meeting in Brussels.

■ International Institute of Communications Forum

On Tuesday 12 and 13th March 2024, KUL participated in the European annual Forum of the International Institute of Communications (IIC). The IIC is an international policy platform for the ICT and digital ecosystem. KUL provided the following talk: AI use in media –recommendations for policymakers. The session focused on the journey towards policy recommendations for the use of AI in the media sector and other results developed in WP2. Many European regulators were in the room and learned about the project and its policy recommendations.

Figure 6: Session 8: AI use in media – recommendations for policymakers

■ AI Café on GenAI

In May 2024, KU Leuven & UvA delivered a talk for the AI-Café which focused on the societal, ethical and legal considerations of the use of GenAI in the media sector. The AI Cafés are an online forum to gain insights into the European AI scene. This Café was the opportunity for the ethical and legal team of AI4Media (KUL and UvA) to share with a wider audience the project's results on these considerations. It constituted an additional opportunity to reach a diverse audience (following the AI Café in general) and present results from different WP2 deliverables in another format. This talk included raising questions about emerging power imbalances in the media landscape, the impact GenAI might have on human creativity, its environmental costs as well as its impact on (dis)information ecosystems and copyright challenges related to harvesting data to build large language models. This Café also reflected on whether and how new EU legal frameworks, such as the AI Act, mitigate these challenges. The Café triggered significant interest with 191 registrations.

The EU Vision for Media Policy in the Era of AI event

On June 19, 2024, the final event of WP2 took place titled “EU Vision for Media Policy in the Era of AI”. Regulators, researchers, practitioners, and even the Belgian Flemish minister of Brussels, Youth, Media and Poverty Reduction, Benjamin Dalle, came together in one room to discuss the transformative potential of AI in the media sector and to look forward to how regulation can make a meaningful impact in how AI is developed and used. The event was hosted at the Belgian Institute for Postal Services and Telecommunications in Brussels and was organised by KUL.

A keynote speech highlighted a two-sided role for policymakers and regulators; regulating challenges and supporting AI developments in Europe. The initial policy recommendations and the findings gathered on our journey towards the final policy recommendations were also shared with the audience. During the event and different panels, it also became clear that one important policy need is to slow down the ongoing flow of new AI legislation, to give regulators and media organisations a chance to implement, enforce and learn from their experiences with the new legal frameworks.

The challenges of developing and integrating AI in the media sector have been discussed by media practitioners including AI4Media use case. It included Rasa Bocyte from the Netherlands Institute for Sound & Vision (NISV), Chaja Libot from the Flemish public broadcaster (VRT), Frank Visser representing the DRAMA project and Angel Spasov from Imagga. While the media practitioners welcomed a pause in the stream of new legislation, they also stressed the importance of keeping up with the new challenges that AI will continue to pose for the media sector, such as the ongoing debate around copy-righted training data. The various panels emphasized the critical need for collaboration across stakeholders active in the media sector. It was highlighted that shaping an effective EU agenda and ensuring the responsible integration of AI in the media sector cannot be achieved in isolation.

The full report about this event has been published online. The event gathered around 65 people, which was the maximum capacity for the venue.

Figure 7: Roundtable Experts' perspective on policy & regulation

4.2 Road towards final Policy Recommendations for AI in the Media Sector

The pilot policy recommendations of D2.4 represented a first attempt to provide guidance on how to better regulate AI in the context of the media sector. They were based on previous deliverables, including the initial whitepaper on the social, economic, and political impact of media AI technologies (D2.2), which also identified specific policy recommendations, as well as a public survey on AI for the media addressed to AI researchers and media industry professionals (D2.3) and a public stakeholders' consultation targeting the research community, media industry and policymakers.

Between 2022 and 2023, three different industry workshops were conducted in the context of WP2 with three different stakeholder groups. The aim of these workshops was to disseminate and qualify the Pilot Policy Recommendations for the use of AI in the Media Sector.

Moreover, three other policy workshops were designed to engage the participants in active discussion over the legal challenges that were highlighted in the initial pilot recommendations and to offer them the chance to provide input on what policies they found to be missing from their vantage point in the sector.

As already mentioned in the introduction to this deliverable, the work conducted through the workshop series led to the project's network extension and enabled the collection of a diverse range of input from stakeholders. The industry workshops enabled to engage with professionals from different sectors and disciplines and geographical regions on recommender systems, content moderation systems and audiovisual archives.

The first policy workshop took place during the Joint Computation + Journalism European Data & Computational Journalism Conference 2023, in Zurich, Switzerland in June 2023, where the participants were representatives from media organizations or media researchers. The second workshop was conducted during the 9th General Assembly of the AI4Media consortium, in Pisa, Italy, in October 2023, where the participants consisted of consortium partners and, as a result, included many technical experts as well as use case partners. The third workshop was conducted with the AI, Media, and Democracy Lab at the University of Amsterdam, Netherlands in November 2023, where the participants included media and legal scholars working in the intersection of AI and media.

The result of the desk research and a series of workshops was a publication of six cross-cutting policy needs for the media sector thoroughly detailed in the 'Final White Paper on the Social, Economic, and Political Impact of Media AI Technologies' published in February 2024.

The publication of the six policy needs coincided with a vivid discussion about the final text of the AI Act. On 13 July 2024, the final version of the AI Act was published. As a result, in the policy recommendations for AI in the media sector we examine these new regulations, particularly in relation to the media sector, identifying areas where the AI Act provides clarity and where it leaves questions unanswered. We offer targeted recommendations to address any gaps. Our focus is on how the AI Act, along with the Digital Services Act (DSA), tackles six key policy needs that are critical for fostering a responsible AI approach within the media industry.

We aim to assist media organizations in understanding the relevant legislation and how it can support their operations. We provide industry professionals with insights and findings they can use in policy discussions, highlighting areas where the legislation may fall short in enabling the media sector to navigate the AI landscape responsibly.

With the AI Act and DSA enforcement just beginning, our policy recommendations are designed to offer interpretive guidance, suggestions, and key considerations for enforcers, national authorities, and media councils, among others.

This section is based on the blog series discussing how the AI Act and the DSA address the needs of the media sector, which were published on Medium and the AI Media Observatory. The series is authored by WP2 humanities partners: Rasa Bocyte from the Netherlands Institute for Sound & Vision, Lidia Dutkiewicz and Noémie Krack, from CiTiP KU Leuven and Anna Schjøtt Hansen from the University of Amsterdam.

The choice to rely on a blog series to communicate our analysis was made to increase the impact and reach of our findings. The blogs are short, appealing and have a flexible format enabling to catch interest and convey our main research takeaways.

The first blog in the series titled 'AI regulation is overlooking the need for third-party transparency in the media sector', was published on July 15, 2024 to coincide with the publication date of the AI Act, and is available here. The second blog titled 'More policies and initiatives need to support responsible AI practices in the media' was published on July 29, 2024

and is available here. The third, titled 'The Big tech-media power imbalance remains despite AI regulation' was published on August 12, 2024 and is available here. The last one, titled 'The AI Act missed an opportunity to address the environmental risks of AI and leaves media to fight the copyright battle' was published on August 26, 2024 and can be found here. In total, the blogposts were viewed over 160 times and read more than 60 times. With time and more dissemination effort, we are expecting these numbers to grow. For the ease of reading, we merged the findings of the blog series below.

4.3 Final Policy Recommendations for AI in the Media Sector

4.3.1 Impose and support good practices for transparency

Transparency is mentioned several times both in the AI Act and the DSA. In the AI Act Recital 27, it is, for example, emphasized that: "Transparency means that AI systems are developed and used in a way that allows appropriate traceability and explainability while making humans aware that they communicate or interact with an AI system, as well as duly informing deployers of the capabilities and limitations of that AI system and affected persons about their rights."

Based on the case study of The Washington Post, researchers Cools and Koliska identified two kinds of algorithmic transparency – internal and external (Cools & Koliska, 2024). The first, internal transparency describes the need to ensure that journalists and other non-technical groups inside media organizations have sufficient knowledge around the AI systems they use. The second, external transparency, refers to transparency practices directed towards the audience to make them aware of the use of AI, such as explanations, clear disclosures, or watermarks (Cools & Koliska, 2024).

Drawing on this distinction, we also identify the third type of algorithmic transparency in media - third-party transparency, which relates to the importance of having insights into how AI systems provided by third-party providers have been trained and how they work.

- Internal transparency

As research shows, the disconnect between increasingly pervasive AI and the level of understanding amongst journalists "could limit journalists' ability to effectively and responsibly use AI systems" (Jones et al., 2022). Closing this intelligibility gap is, therefore, necessary for journalists and media professionals in general to also question and challenge AI outputs, and to have a sufficient level of understanding to report responsibly on AI and algorithms in society. The need for improving the intelligibility of AI and fostering AI literacy is also recognized by the AI Act, which highlights how providers and deployers of AI systems including media organizations must ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with the operations and use of AI systems on their behalf (Art. 4 AI Act). The context the AI systems are to be used in as well as the persons or groups of persons on whom the AI systems are to be used should be considered. However, as highlighted by DW Innovation, implementing sufficient governance and supporting literacy is highly resource-

intensive and takes a lot of translational work and, so far, the regulation remains vague on how media organizations will be supported in this work (DW Innovation 2024). Thereby, leaving the burden on the individual organizations.

- External transparency

The AI Act requires providers of AI systems, such as OpenAI, to make it clear to users that they are interacting with an AI system, unless this is obvious from the circumstances and the context of use (Art. 50(1)).[6] This means that providers should, at least, explicitly inform all the natural persons interacting with those AI systems (including deployers) about the artificial nature of the entity they're interacting with. "In that way the risk of deception of, or anthropomorphization by a natural person should be reduced" (Gils, 2024).

As a rule, providers of AI systems that are capable of generating synthetic audio (e.g. voice generation such as Amazon's Alexa), image (e.g. Dall-E, Midjourney), video (e.g. OpenAI's Sora) or text (e.g. OpenAI's ChatGPT) content, must ensure that such content is marked in a machine-readable format and detectable as AI-generated or manipulated (Art. 50(2)). One of the exceptions apply if the AI systems perform an assistive function for standard editing or do not substantially alter the input data provided by the deployer or the semantics thereof.

For now, it remains unclear what forms of transparency will be sufficient and whether they will be meaningful to the audience. The different techniques exist, and each type of output requires a different approach. As noticed by Gils, currently available (water)marking solutions and related detection tools often still suffer from various limitations and drawbacks (Gils, 2024). The lack of interoperability between marking solutions and detection tools developed by separate providers is another important issue. Adoption of respective standards such as the Coalition for Content Provenance and Authenticity (C2PA)-standard[7] should facilitate compliance with Article 50(2) AI Act. It is also unclear who will benefit from this marking obligation. As Gils asks, "Should only parties that have the ability to develop or acquire detection tools be able to detect content as artificially generated or manipulated? Or should all possible stakeholders be able to do so? If so, how can it be ensured that exposed natural persons are aware of how to detect the applied marking?" (Gils, 2024). These are just some of the many open questions.

Another relevant transparency provision for the media sector is Article 50(4) AI Act. Article 50(4) para 1 AI Act requires deployers of an AI system (i.e. media organizations) that generates or manipulates image, audio or video content constituting a deep fake, to disclose that the deep fake content has been artificially generated or manipulated. Article 3(60) AI Act defines this as

[6] What "interaction" means in this context is unclear, but it could cover applications such as chatbots, newsbots, recommender systems, and automated writing systems. See: Helberger and Diakopoulos, "The European AI Act and How It Matters for Research into AI in Media and Journalism" (2022) Digital Journalism.

[7] See the Coalition for Content Provenance and Authenticity's website: https://c2pa.org/faq/ and https://c2pa.org/specifications/specifications/2.0/index.html;

"AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful".[8]

In theory, disclosure through labelling is presumed to empower readers to critically evaluate information and make informed decisions, mitigating concerns around misinformation and misleading content (Zier, 2024). Yet, it remains unclear what forms of transparency will be sufficient and whether they will be meaningful to the audience or risk becoming another pro-forma transparency practice, such as the 'cookie consent form'.

Moving to the media sector specifically, the Nordic AI Journalism network has provided some initial guidelines directed at media on when and how to be transparent to the audience (Nordic AI Journalism 2024). One of the recommendations is to be specific about the type of AI tool applied. At the same time, technical details (such as model versions and model weights) may be "redundant and may even be counterproductive by making the consumer feel they do not understand the technical jargon" (Nordic AI Journalism 2024). A relevant exception may apply to method descriptions of editorial work, especially investigative journalism, where AI has a major journalistic impact.

Another recommendation by the Nordic AI Journalism network is to encourage harmonization. Harmonised or (even standardised) labelling could be one of the solutions. As a first step, the network recommends a harmonized language regarding generative AI: "As a basis, we suggest the wording "created with the support of", to signal AI's actual impact on the content and remind the media consumer that there is an editorial process (and staff) behind the content. Alternative wordings that have been discussed but dismissed include "assisted by," "together with," "via," or "by." We believe these terms set incorrect expectations and may diminish the perceived role of humans in the editorial process."

The network also advises against visual labelling (e.g., an icon) for AI in editorial media. This observation can be extended beyond the Swedish media industry. Drawing on an example of Hoodline article containing an AI label (see Figure 8), Zier asks: "Would a simple tag saying "AI" suffice as a meaningful transparency approach?" (Zier, 2024).

[8] See e.g Gils 2024, asking: "Does this entail that the deep fake concept only applies when the AI-generated content depicts e.g. current events/affairs or currently living persons, excluding historical events and figures?" What *"falsely appears as authentic or truthful"* to one person may be evidently fake to another.

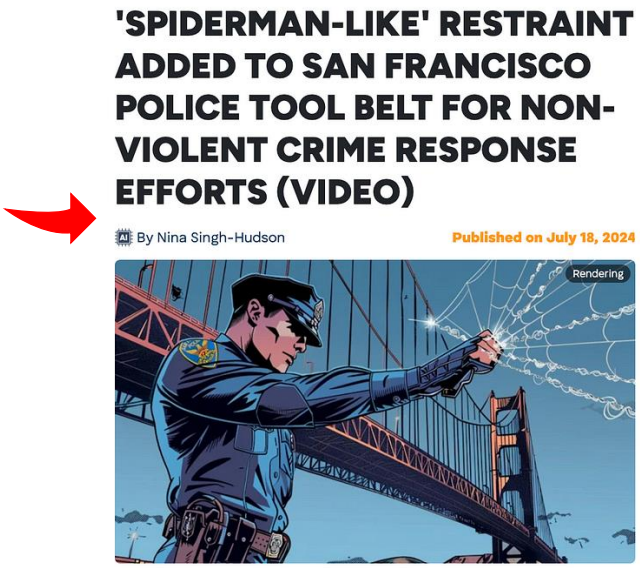


Figure 8: Hoodline's "AI" label appearing next to fake author personas on the site has been criticized for being an empty gesture

How the exact label design, its placement, size, etc. should look like to be effective is therefore a key question that will need to be tackled – and hopefully agreed on – by the media sector. As research shows, the use of terms such as 'artificially generated' or 'manipulated' for labelling content is differently perceived by the audience. In particular, 'manipulated' is understood as a term for labelling content that is misleading (Epstein et al., 2023).

Going back to Art. 50 AI Act, importantly, there are some exceptions which are relevant to the media sector. First, if the deep fake content forms part of an evidently artistic, creative, satirical, fictional or analogous work or programme, the disclosure only has to happen in an appropriate manner that does not hamper the display or enjoyment of the work. Second, deployers of an AI-generated or manipulated text which is published with the purpose of informing the public on matters of public interest shall disclose that the text has been artificially generated or manipulated. Yet, if the AI-generated text has undergone a process of human review or editorial control within an organisation that holds editorial responsibility for the content (such as a publisher), disclosure is no longer necessary.

While these exemptions will make it easier for media to comply with the AI Act, they also produce potential issues. First, this provision raises questions as to what will count as a human review or editorial control and who can be said to hold editorial responsibility. Whether AI-generated content has been subject to these controls will be up to the regulator to assess— a highly difficult task.

Moreover, as research shows, a large majority of respondents want media organizations to be transparent and provide labels when using AI (Fletcher & Nielsen, 2024). "Asked whether news that has been produced mostly by AI with some human oversight should be labelled as such, the vast majority of respondents want at least some disclosure or labelling. Only 5% of our respondents say none of the use cases we listed need to be disclosed." says the recent study (Fletcher & Nielsen, 2024). Although there is less consensus on what uses should be disclosed

or labelled, around half of respondents think that using AI for 'writing the text of an article' should be disclosed (47%). With the already growing mistrust in media sources, it will be even more important that media organizations contribute to ensuring trust via transparency practices, which the current exemptions could at a minimum de-incentivize.

Turning to the DSA, the text also provides some provisions around external transparency, including that online platforms must publish their terms and conditions (T&C) in easily understandable language and openly report what content is moderated. In their T&C they must also include a description of the tools used for content moderation, including AI systems that either automate or support content moderation practices. In addition, users must be able to report harmful content and also receive a statement clarifying why the content was moderated. In case of a complaint, it is also required that a human must be in the loop.

To support transparency around these systems, these platforms must also draft yearly transparency reports on content moderation including a qualitative description, a specification of the precise purposes, indicators of the accuracy and the possible rate of error of the automated means used. In practice, this means that users are in a better position to understand why and how their content, for example, was removed, which might be important for media organizations to contest potential restrictions on their content (see Article 14–17 DSA).

The European Media Freedom Act (EMFA)[9] similarly requires very large online platforms (VLOPs, see note 10) to annually publish information on the grounds for any number of times they restricted or suspended access to the content of media service providers (see Art. 18 EMFA). It also requires VLOPs before suspending or restricting taking effect to communicate to the media service provider concerned a statement of reasons and to give a 24-hour window to reply.

- Third-party transparency

In both the DSA and AI Act, there are no provisions that make information about how the model was trained and how it works clearly available for the buyer. In the AI Act, only for general-purpose AI models (such as GPT3) or high-risk AI systems, there are requirements for providing some information about the training datasets and documentation around the capabilities and limitations of the models (see Recital 66 & 67, Article 53, and Annex XII).

Beyond the DSA and AI Act, the Council of Europe has recently published 'Guidelines on the responsible implementation of artificial intelligence (AI) systems in journalism'. The Guidelines highlight how responsible procurement requires insights into the systems that are purchased — something that is currently not strongly stipulated in any regulation, which minimizes the bargaining power of media. The report includes a checklist for media organizations to guide the procurement process, which lists several central themes and questions that could help in

[9] Regulation (EU) 2024/1083 of the European Parliament and of the Council of 11 April 2024 establishing a common framework for media services in the internal market and amending Directive 2010/13/EU (European Media Freedom Act), OJ L, 2024/1083, 17.4.2024.

assessing the suitability of a particular AI provider, and in scrutinizing the fairness of a procurement contract with an external provider.

■ Recommendations

Figure 9 offers recommendations for good practices for AI transparency.

Our recommendation

- We encourage strategies to support the closure of the intelligibility gap in the media sector and argue for sustainable funding to support the AI literacy of media staff.
- We encourage the AI Office to draw a code of practice to facilitate the effective implementation of the obligations regarding the detection and labeling of AI-generated or manipulated content under Article 50 (7) AI Act.
- The AI Office (and the Member States) could also use its mandate in Article 95 AI Act to facilitate the drawing up of codes of conduct concerning the voluntary application of specific requirements of the AI Act – such codes of conduct could relate to best AI transparency practices.
- Article 96 (1)(d) AI Act obliges the European Commission to develop guidelines on the practical implementation of the transparency obligations laid down in Article 50 AI Act. When issuing such guidelines, the European Commission should pay particular attention to the needs of small and medium-sized enterprises including start-ups, of local public authorities, and of the sectors most likely to be affected by the obligations of Article 50 AI Act. We encourage the EC to be in dialogue with the media stakeholders to take into account the needs of the media sector.
- In the lack of harmonized standards and common specifications on labeling, we encourage media organizations, associations, and relevant media authorities to work together towards a common approach to labeling the use of AI in the news (and media) sector.
- We encourage policies that require increased transparency of third-party providers of AI solutions. We opt for more support and guidance for media organizations to guide the procurement process and for scrutiny of whether the sellers of the AI solutions share the relevant information.

Figure 9: Recommendations towards good practices for AI transparency

4.3.2 Support research in and of AI solutions and stimulate responsible development of AI

■ Access to datasets and system APIs for research and investigative purposes

Access to datasets and Application Programming Interface (APIs) for research has been addressed in Article 40 of the DSA which requires VLOPS and VLOSEs to give access to their datasets and APIs for research purposes. While it's a great step forward, the well-intended provision has received some criticism. First, it requires that researchers are 'vetted', meaning

they have to be affiliated to a research organization. Although Article 40(12) also mentions "those affiliated to not-for-profit bodies, organizations and associations", the vetting criteria risk leaving out researchers working in the media sector, for example, investigative journalists. Since the European Commission Delegated Act has yet to be published in its final form, and much feedback was received there is hope that the scope of 'vetted researcher' could be extended to include other actors. Moreover, the provision does not address access for training purposes, which can seriously hinder the research into AI solutions that, for example, address important language gaps. Second, the procedure to become vetted is complex and burdensome. VLOPs can also exploit this to their advantage and be selective about access conditions they set out and what type of datasets they make available. At this point in time, we are still waiting for a delegated act to be adopted which will lay down the specific conditions under which VLOPs and VLOSEs are to provide data. Third, these access requests are only eligible for research centered on understanding the systemic risks posed by the platforms and their risk mitigation measures. This, together with the fact that the researcher must already demonstrate that the 'expected results of that research' will contribute to an analysis of systemic risk, could exclude exploratory research aimed at identifying new risks and formulating new research hypotheses.

It is too early to assess the effectiveness of Art. 40 DSA. And while we are waiting for the delegated act, the platforms continue their '*business as usual*' strategy: some limit the eligibility of researchers based on their geographic location, other ask that researchers share with them excessive details on their research projects or require researchers to agree to platforms' terms whereby they must provide a copy of their research output before publication (Carvalho, 2024).

- Organisational upskilling and guidelines on responsible practices

Neither the DSA nor the AI Act have any direct programs oriented towards supporting upskilling in the sector or providing solutions to the skewed labor market.[10] The DSA does, however, expect online platforms to "guarantee sufficient human and financial resources" (recital 43) and VLOPs to have sufficient human resources, including the content moderation personnel, their training, and local expertise (recital 87). The Digital Services Coordinators (i.e. national authorities) shall have the "necessary number of staff and experts with specialized skills" (recital 111, Art. 50).

Similarly, Article 4 in the AI Act requires providers and deployers of AI systems to ensure, to their best extent, a sufficient level of AI literacy of their staff and other persons dealing with AI systems on their behalf. However, exactly what that entails, and what 'extent' of expertise is expected from media organizations using AI, is not clear.

[10] However, the EU has launched programs and initiatives, such as the AI Skills Strategy for Europe, ARISA — AI Skills Needs Analysis and media-specific calls such as 'Fostering European Media Talents and Skills' and the business cluster under the MEDIA strand of Creative Europe which aims to promote business innovation, scalability and talents across the European audiovisual industry's value chain.

- Support the production of open and shareable datasets

As the EC study on Improving access to and reuse of research results, publications and data for scientific purposes shows, researchers and research organizations have to bear the burden of complexity and legal uncertainty in data access and reuse for research purposes. Despite the importance of producing and sharing responsible datasets, the current legal landscape remains very fragmented and complex and neither the AI Act nor the DSA addresses this policy need directly. Some even argue that "trade secrecy" around algorithms will be a significant barrier to effectively enforcing the DSA (Miller, 2024). Outside the direct legislative landscape, we do see great potential in the EU-funded initiative Common European Data Spaces, which could make it possible to create shareable datasets. In the media sector, TEMS (Trusted European Media Data Space) is a cross-border project tasked to build and run a data space for the media industry. This backbone infrastructure could provide a potential avenue for supporting the needs of the media sector.

Another avenue towards open data sets is open source. Open source AI refers to the use of open source AI components (e.g. AI model documentation, training data) that are under Open Source (OS) licenses, i.e. licenses that comply with the open source definition (Benhamou, 2024). Open source companies generally disclose more information about the AI system's architecture, datasets, and training methods, making more data and documentation available. However, it has been argued that models that claim to be open source, such as Llama 2, Falcon, or Mistral, are being released without providing access to training datasets or even basic information about them (the so-called 'openwashing') (Tarkowski 2024). Due to the limited transparency requirements in the AI Act regarding datasets used for training open-source models, it is not expected that this particular legislation will support the creation of a marketplace of ethical open-source datasets.

- Sustainable funding schemes & public-private collaborations

AI innovation in the media sector — in particular in public organizations — often relies on competition-based financing for projects. While these project-based innovation actions are essential to enable research and concrete implementation of AI solutions in the media sector, they should be guided by a long-term strategy that prioritizes sustainability. More generally, the media sector could benefit from more sustainable funding and support mechanisms with a long-term vision, to also work as a counter mechanism to the short-term funding provided by a variety of commercial actors, such as Google's Digital News Innovation Fund (DNI).

If we look at the different EU funding mechanisms available for the media sector, they are currently almost exclusively project-based. This means that it is up to each organization to figure out how they will fund the maintenance and long-term operation of AI services they develop through these projects. Furthermore, it becomes the organization's responsibility to determine how their projects can contribute to the broader ecosystem of AI developments in media.

There are some emerging initiatives, such as the new AI innovation package and the introduction of regulatory sandboxes that could begin to address the need for sustainable innovation funding and continue to stress the importance of private-public collaborations. Similarly, MediaInvest is the European Commission's equity investment instrument that helps to bridge the financial gap in the audiovisual sector by stimulating more investment. However, these also have short-term horizons and they unfortunately do not have specific provisions for the media sector. The GenAI4EU initiative funded under the AI innovation package will, for example, support startups and SMEs in 14 industrial ecosystems, but media is not one of them with the exception of public service media which will fall into the category with all public sectors.

■ Recommendations

Figure 10 offers recommendations for supporting research in and of AI solutions and stimulating the responsible development of AI.

Our recommendation

- We encourage the EC to clarify in its Delegated Act on data access the exact scope, conditions, and mechanisms for data access for researchers, including also (investigative) journalists.
- We support policies promoting publicly available datasets and open data access and open APIs for research or investigative purposes (e.g., training data or algorithmic accountability reporting) which go beyond DSA's 'systemic risk'.
- We encourage relevant authorities to stay vigilant about 'open washing' and to make sure that data access gatekeepers do not circumvent the applicable provisions and provide meaningful insights into the workings of the AI model.
- We encourage project funders to support and fund data spaces for the media industry, the development of open datasets and AI solutions for low-resource languages.
- We call on funders to focus on long-term funding schemes for the benefit of society that prioritize sustainability over short-term quick results.

Figure 10: Recommendations for supporting research in and of AI solutions and stimulating responsible development of AI

4.3.3 The need to mitigate AI divides and power imbalances

To mitigate the growing AI divide, we found that there is a need for (i) supporting organizational capacity building particularly for small organizations and those situated in geographical regions with historical disadvantages; (ii) increased local and regional collaborations to limit the costs of developing AI; and (iii) the development of open datasets and AI solutions for minority languages to counteract the fact that most AI solutions are developed for and work best in English.

- Regulatory sandboxes

In the AI Act, there are some interesting provisions when it comes to the AI sandboxes (Article 57 & 58) that could in part help support organizational capacity building and regional collaboration, as they are aimed to "encourage collaboration among AI ecosystem actors, facilitating cooperation between public and private sectors". These "safe space" environments to test new innovative products and services are supposed to enable broad and equal access to AI development opportunities and are free for SMEs and startups (Yordanova 2024). Thereby, providing a way for small or low-resource actors to participate in the AI landscape. In addition to technology development, the sandboxes will enable participants to receive guidance, supervision and support within the sandbox with a view to identifying risks, in particular to fundamental rights and receive guidance on regulatory expectations and how to fulfill the requirements and AI Act obligations.

- Open source models and AI solutions for minority languages

In February 2024, the EC established both the European Common Media Dataspaces and the European Digital Infrastructure Consortium for the Alliance for Language Technologies (ALT-EDIC). The ALT-EDIC aims to develop a common European infrastructure in language technologies to address the shortage of European language data for the training of AI solutions. This will among other results in the creation of 'Large Language Models' of European regional and official languages. While this is not legislation per se, it still offers an important avenue for the development of open datasets and AI solutions for minority languages.

Open source of both models and datasets is also often discussed as an avenue for minimizing growing disparities in access to AI and its benefits. In the AI Act, steps are also made to incentivize open-sourcing models and datasets, such as exempting open-source AI systems that have limited or minimal risk from any legal obligations under the Act. However, the providers of these models are currently not legally required to provide technical documentation around these models (unless they pose a systemic risk) or to make training datasets available, which makes it difficult to use these models responsibly due to a lack of insight into how the models function (Tarkowski 2024). What openness means and how it can be ensured in the context of open source AI development is, therefore, not directly addressed by the AI Act, and questions of access to, for example, training datasets will therefore depend on good faith and self-regulation.

- Alternative funding and independent AI infrastructure

To address the growing power imbalances in the AI landscape, we found that there is a need to support (i) alternative funding schemes that allow organizations developing AI to be independent of funding from big tech companies; (ii) the development of public and open-source critical AI infrastructures and models (e.g., LLMs) to ensure a diverse landscape of AI, and (iii) the need to enable organizations to have more negotiation and bargaining power with third-

party providers. Currently, there are limited attempts to address the growing power imbalances in the AI landscape between the large model providers and their users and the 'infrastructure capture' (Nechushtai 2017) in the EU legislation. What can be seen is mainly in relation to foundation models where the transparency obligations on training data for foundation models is a step towards empowering right holders in their licensing and partnership negotiations with big tech companies. Equally, the decision to hardly regulate smaller models was very much such a step to enable a power shift towards the so-called "European Champions".

However, a few other emerging policies do try to address the issue, such as the Digital Markets Act (DMA)[11], which imposes obligations and rules for large online platforms ("gatekeepers"), the Data Act[12], which aims to foster competitive data markets, and the EMFA that focuses on empowering media and protecting editorial independence. It remains to be seen how these regulations are enforced and whether they will enable a digital level playing field and support the counterbalances in the digital market.

The EU is also taking steps to provide alternative infrastructures than those offered by big tech, for example, via their 'AI Factories' that aim to unite and leverage the network of supercomputers in Europe. However, the investments made by the EU in infrastructures remain microscopic in comparison to those made by, for example, Microsoft and they are also often less user-friendly. So even when offered for free to, for example, EU-based startups, they chose the established actors which was seen, for example, with the start-up Mistral AI.[13]

- Recommendations

Figure 11 offers recommendations to mitigate AI divides and power imbalances.

[11] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) PE/17/2022/REV/1, OJ L 265, 12.10.2022, p. 1–66.
[12] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act) PE/49/2023/REV/1, OJ L, 2023/2854, 22.12.2023.
[13] https://www.reuters.com/technology/microsofts-deal-with-mistral-ai-faces-eu-scrutiny-2024-02-27/

Our recommendation

- We encourage national authorities to establish regulatory sandboxes to enable testing and experimentation for media organizations.
- We encourage the initiative to create LLMs in European regional and minority languages.
- We encourage the creation of the standardized frameworks for model openness assessment made up of multiple elements and we encourage national authorities to invest in the funding schemes that can help enable regional and local collaborations on AI that can allow for the development of scalable and open-source AI solutions to be used in multiple organizations.
- We encourage alternative funding schemes that allow organizations developing AI to be independent of funding from big tech companies.
- We suggest strategies which support media organizations in AI integration and the development of localized solutions, independently from big tech providers.

Figure 11: Recommendations to mitigate AI divides and power imbalances

4.3.4 The need for global and societally focused policies

■ A globally focused regulation

Many media organizations and AI developers find it difficult to navigate the AI legal landscape and also fear that it will not have the desired effects due to the globalized nature of AI. However, EU is tied by its limited (to EU Member States) regulatory competencies. Yet, the AI Act is expected to become a global standard through the "Brussels Effect."[14] Some researchers also argue that the EU's AI Act will incentivize changes in products offered by US tech companies (especially "high-risk"), as well as influence regulation adopted by other jurisdictions (Siegmann and Anderljung 2024). On the other hand, since the AI Act follows EU product safety legislation, its provisions *"supply limited protection to some of the values the EU policy intends to protect, such as the protection of fundamental rights"* (Almada and Radu 2024). The global diffusion of the AI safety standards needs to be distinguished from the diffusion of regulatory goals and framings, which might not accompany, or even be undermined by, the former under the Brussels Effect. (Almada and Radu 2024).

[14] 'Brussels Effect' refers to "EU's unilateral ability to regulate global markets by setting the standards in competition policy, environmental protection, food safety, the protection of privacy, or the regulation of hate speech in social media" see: Anu Bradford, The European Union in a globalised world: the "Brussels effect", Aug 2021, 75-79 https://geopolitique.eu/en/articles/the-european-union-in-a-globalised-world-the-brussels-effect/.

Outside the EU, the Committee of Ministers of the Council of Europe have recently adopted the Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law,[15] which aims to establish a global minimum standard for protecting human rights from risks posed by AI. In contrast to the EU AI Act, the Convention is an international, potentially global treaty with contributions from various stakeholders, including the 46 Council of Europe member States, and a number of countries from several regions (Argentina, Australia, Canada, Costa Rica, the Holy See, Israel, Japan, Mexico, Peru, Uruguay and the United States).

The Convention will be opened for signature by the States on 5 September 2024. It is up to each signatory State to decide on the measures to comply with the Convention's provisions. Despite the global reach, the Convention is not without its criticism, as voices have raised concern about the open door the convention leaves open for the private sector (Reporters Without Borders 2024) and that states can continue to allow AI to be developed and used under the guise of "national security" in cases such as face recognition in public spaces, AI border protection, or scanning social media profiles (Algorithm Watch 2024).

- A societally focused legislation

While it remains crucial that the emerging AI legislation protects individual risks and rights, it, unfortunately, has led to a limited focus on accounting for societal risks produced by AI, such as the potential for increased polarization in a time of growing AI-generated misinformation, the risk of new displacement patterns across the labor force and the environmental risks posed by AI.

In Article 1 of the AI Act, it is also described how the purpose of improving the functioning of the internal market should take place "while ensuring a high level of protection of health, safety, fundamental rights, including (...) environmental protection." Environmental and societal aspects of AI are mentioned several times in the Recitals but do not translate into concrete, legally binding obligations.

Generally, the EU is 'trapped' by the legal basis of the treaty that focuses only on the internal market, which leads to legal instruments being strongly market-oriented, as opposed to societally oriented. However, there are some small examples of how the current legislation does address some of these more societally oriented policy needs.

As an example, Recital 27 points out that "social and environmental well-being means that AI systems are developed and used in a sustainable and environmentally friendly manner as well as in a way to benefit all human beings while monitoring and assessing the long-term impacts on the individual, society and democracy. The application of those principles *should be translated, when possible,* in the design and use of AI models." Equally, Recital 142 mentions

[15] Council of Europe Treaty Series - No. [225] Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law [Vilnius, 5.IX.2024].

that "to ensure that AI leads to socially and environmentally beneficial outcomes, Member States *are encouraged* to support and promote research and development of AI solutions in support of socially and environmentally beneficial outcomes, such as AI-based solutions to increase accessibility for persons with disabilities, tackle socio-economic inequalities, or meet environmental targets, by allocating sufficient resources, including public and Union funding (...). Such projects should be based on the principle of interdisciplinary cooperation between AI developers, experts on inequality and non-discrimination, accessibility, consumer, environmental, and digital rights, as well as academics."

- Worker displacement

Diving more into the specific societal concerns, worker displacement has been considered a major risk both earlier and particularly with the rise of generative AI. In an AI Now Institute report, they highlight the need to put in place mechanisms that can address the rise of new displacement patterns in the workforce (AI Now 2019). In general, due to the EU remit, in the context of employment and workers' protection, the AI Act does not affect EU social policy or national labor law (Recital 9). There are, however, ways to interpret the category of harm or societal impact, including work displacement issues.

In Article 27 of the AI Act, there is an obligation for a Fundamental Right Impact Assessment for high-risk AI systems. The assessment should include who is likely to be affected (including which category or group of people) and what the risks of harm impacting them are. This could be interpreted as an obligation for deployers of AI systems to account for potential new worker displacements caused by the deployment and ensure there are mechanisms in place to protect those workers.

- Copyright and intellectual property

The question of copyright has been a heated one since the release of ChatGPT, followed by a range of lawsuits claiming that the training data consists of copyrighted data.[16] This concern is very evident in the media sector whose intellectual property (text and images) currently is at risk of being abused - but here the AI Act does afford some protection.

[16] See e.g. *8 Daily Newspapers Sue OpenAI and Microsoft Over A.I.,* https://www.nytimes.com/2024/04/30/business/media/newspapers-sued-microsoft-openai.html; *Google sued by US artists over AI image generator*, https://www.reuters.com/legal/litigation/google-sued-by-us-artists-over-ai-image-generator-2024-04-29/. For an overview, see also AI4Media repository: https://docs.google.com/spreadsheets/d/1UZF-H-SfzpeaZ7OjEmbJFippDESyRilfX3l5l_US7uE/edit?usp=sharing.

While the AI Act does not include new provisions on exceptions to the copyright, it does remind us that the Directive on Copyright in the Digital Single Market[17] requires that any use of copyright-protected content requires authorization unless relevant copyright exceptions and limitations apply (Recital 105).

Article 53(1)(c) states that providers of general-purpose AI models should put in place a policy to identify and comply with a reservation of rights (a so-called 'opt-out'). Yet, this provision does not offer any further guidelines for content creators and rightsholders. The question of how the opt-outs should work in practice is largely unclear, as there are currently no generally recognized standards or protocols for the machine-readable expression of the rights reservation (Keller 2023).

Article 53(1)(d) also requires that providers will have a publicly available detailed summary of the content used for training of the general-purpose AI model - including any copyrighted material. However, this still only provides an opt-out option moving forward, leaving the status of already scraped content unclear even if it is described in such summaries. Recital 107 further specifies that it should be generally comprehensive to facilitate parties with legitimate interests, including copyright holders, to exercise and enforce their rights under EU law. This could include, for example, listing the main data collections or data sets that have been used for the model training, such as large private or public databases or data archives, and providing a narrative explanation about other data sources used. More details on what the "sufficiently detailed summary" means should be specified in a standard template provided by the AI Office.

AI4Media established a repository to observe how media organizations are responding to content scraping activities that fuel AI models. It can be accessed here. The repository aims to guide media creators and holders as they formulate their own responses and seek solutions to these developments. It also provides researchers and policy-makers with insights into policy gaps and challenges that media makers face against tech companies (Bocyte and Dutkiewicz 2024).

In the repository one can find content related to the following categories:

- Anti-scraping statements — positions expressed by media creators and holders against automated crawling of their content;
- Legal action — lawsuits launched against tech companies for unlawful data scraping;
- Licensing deals — agreements between media holders and tech companies to use media content for model training;
- New techniques & methods — tools that are being developed to prevent or detect content scraping;
- Opinion pieces and analysis — general analysis of the topic.

[17] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC, PE/51/2019/REV/1, OJ L 130, 17.5.2019, p. 92–125.

■ Environmental risks

Last, is the question of environmental risks. While AI can be an important tool in efficiently solving environmental problems that societies face, the environmental impact of the development and (re)training of AI models cannot be understated. Initially there was a proposal to include a measurement of reasonably foreseeable adverse impacts of putting the system into use as an element of fundamental rights impact assessments for high-risk AI systems. However, it did not make it to the final text. As a result, the EU AI Act's approach to environmental issues has been considered "a missed opportunity to make the AI industry more environmentally sustainable." (Laranjeira de Pereira 2024).

That said, there are small steps in place. Article 40, does state that the EU will develop standards for high-risk AI systems that will be aimed at reducing the system's consumption of energy and other resources during its lifecycle and that will require energy-efficient development of general-purpose AI models. Equally, Article 51 combined with Annex XIII states that the estimated cost of training, time required for the training, or energy consumption for the training, will be one of the criteria to take into account to categories a model as a general-purpose AI model.

On the softer side, Article 95 notes that Member States are tasked with supporting the development of codes of conduct for voluntary adherence by deployers, aiming to address specific requirements for all AI systems - including assessing and minimizing the environmental impact of AI systems and promoting energy-efficient programming and techniques throughout their lifecycle.

Importantly, Article 112 in the AI Act leaves a door open to review and a stricter approach when it comes to the energy and environmental costs of AI further down the line.

Recommendations

Figure 12 offers recommendations to support global and societally focused policies.

Our recommendation

- We encourage the development of opt-out compliance policies and an establishment of generally recognized standards or protocols for the machine-readable expression of the rights reservation and opt-outs from training generative AI models.
- We encourage initiatives which focus on the collective responses of media content creators to data scraping for model training.
- We encourage the AI Office to adopt a standard template on what the "sufficiently detailed summary" provided by general-purpose AI (GPAI) developers means; such template should in particular contain elements regarding the ethical review process of the training data, its collection, legal basis, its diversity and whether the data was scraped from the internet and information about the crawling method.
- We encourage the Member States to use their mandate in Article 95 AI Act to develop codes of conduct for the assessment and minimization of the environmental impact of AI systems.

Figure 12: Recommendations to support global and societally focused policies

5 Conclusions

D2.6 is the final deliverable of WP2 “European AI Vision, Policy and Common Research Agendas”. WP2 involved many activities, deliverables, events, and workshops, focusing on the use of AI in media and the policy and legal landscape.

The main outcomes presented in D2.6 include the updated version of the AI4Media Strategic Research Agenda and the final policy recommendations on AI for the media sector.

The AI4Media SRA focuses on the use of AI technology in the service of the media, society and democracy. The first version of the Agenda presented four main research pillars focused on developing novel learning paradigms, trustworthy AI methods, multimedia content analysis and production tools, and human and society-centered AI applications for the media industry. The second version introduces a new research pillar, namely Generative AI and LLMs, exploring recent developments and relevant challenges, and offering insights on emerging research trends and their potential impact.

The SRA aims to serve a useful point of reference for the AI research community, the media industry and policymakers. It should be seen as a snapshot of the current situation with regard to AI technology for the media industry and the research activities of the consortium. As AI and its applications for the media are advancing fast, new high-impact research areas and applications will be unveiled that will need to be explored in possible future versions of this report.

With regard to **policy recommendations**, it is clear from the analysis offered in this deliverable that there are numerous research directions to explore in the domain of AI, GenAI and their ethical, societal and legal challenges. Whereas some have a more general focus, others, such as copyright protection of media organizations and content creators' content, are more media sector-oriented. Undoubtedly, for the better future of AI and media, it is of paramount importance that adequate research resources and funding goes towards the exploration of these research directions.

Moreover, our analysis showed that the EU has now developed its robust legal framework regulating AI (including the AI Act, the DSA). Some of the challenges of the initial Pilot Policy recommendations for the use of AI in the media sector have therefore been tackled. Yet, some of the gaps remain as shown in this deliverable. Some relate to further regulatory action: the adoption of the Delegated Act for data access, codes of practice for labelling of AI-generated or manipulated outputs, guidelines and codes of conduct to govern some of the environmental risks posed by AI. Others will need to be tackled by self-regulation, and practical and technical means (e.g. on the copyright protection and reservation of rights).

Importantly, at the event ‘*EU Vision for Media Policy in the Era of* AI’ organized by KUL in June 2024, one of the key takeaways was also the need to slow down the development of new legislation and start focusing on enforcement to get a better sense of the actual impact the AI

Act, and other recent legislative instruments might have. So, while we have identified a series of gaps and offered recommendations of how to tackle them, we must also be patient and first see what effect the many good tenets of the legislation will have.

Hopefully, even going beyond the timeframe of the AI4Media project, the analysis of the developments of AI in media, the enforcement and interpretation of the newly adopted EU legal frameworks will continue as part of the AI Media Observatory.

6 References

AI4Media, 2024 *It is time to enforce AI regulation before adding more to the mix* https://www.ai4media.eu/it-is-time-to-enforce-ai-regulation-before-adding-more-to-the-mix/

Algorithm Watch, 2024, *The Council of Europe's Convention on AI: No free ride for tech companies and security authorities!* https://algorithmwatch.org/en/council-of-europe-ai-convention/

Almada M., and Radu A., 2024 *The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy | German Law Journal | Cambridge Core*. (n.d.). Retrieved 7 August 2024, from https://www.cambridge.org/core/journals/german-law-journal/article/brussels-sideeffect-how-the-ai-act-can-reduce-the-global-reach-of-eu-policy/032C72AEC537EBB6AE96C0FD90387E3E#fn160

Benhamou, Y. (2024, April 15). *Open Source AI – definition and selected legal challenges*. Kluwer Copyright Blog. https://copyrightblog.kluweriplaw.com/2024/04/15/open-source-ai-definition-and-selected-legal-challenges/

Bocyte, R., Krack, N., Dutkiewicz, L., Schjøtt Hansen, A., 2024 *Blog series: More policies and initiatives need to support responsible AI practices in the media* https://medium.com/ai-media-observatory/blog-series-more-policies-and-initiatives-need-to-support-responsible-ai-practices-in-the-media-2a42d271d1e1

Bocyte, R., Dutkiewicz, L. (2024) *How is the Media Sector Responding to Content Crawling for Model Training* https://medium.com/ai-media-observatory/how-is-the-media-sector-responding-to-content-crawling-for-model-training-9812ac2916d8

Carvalho de, M. C. (2024, May 31). *Researcher Access to Platform Data and the DSA: One Step Forward, Three Steps Back* | Tech Policy Press. https://techpolicy.press/researcher-access-to-platform-data-and-the-dsa-one-step-forward-three-steps-back

Cools, H., & Koliska, M. (2024). News Automation and Algorithmic Transparency in the Newsroom: The Case of the Washington Post. *Journalism Studies*, *25*(6), 662–680. https://doi.org/10.1080/1461670X.2024.2326636

Crawford, Kate, Roel Dobbe, Theodora Dryer, Genevieve Fried, Ben Green, Elizabeth Kaziunas, Amba Kak, Varoon Mathur, Erin McElroy, Andrea Nill Sánchez, Deborah Raji, Joy Lisi Rankin, Rashida Richardson, Jason Schultz, Sarah Myers West, and Meredith Whittaker. AI Now 2019 Report. New York: AI Now Institute, 2019, https://ainowinstitute.org/AI_Now_2019_Report.html.

DW Innovation*, AI in Media Tools: How to Increase User Trust and Support AI Governance*. (2024). Retrieved 7 August 2024, from https://innovation.dw.com/articles/ai-media-tools-user-trust

Epstein, Z., Fang, M. C., Arechar, A. A., & Rand, D. (2023). *What label should be applied to content produced by generative AI?* OSF. https://doi.org/10.31234/osf.io/v4mfz

Fletcher, R., & Nielsen, R. K. (2024). *What does the public in six countries think of generative AI in news?* Reuters Institute for the Study of Journalism. https://reutersinstitute.politics.ox.ac.uk/what-does-public-six-countries-think-generative-ai-news#header--6

Gils, T. (2024). *A Detailed Analysis of Article 50 of the EU's Artificial Intelligence Act*. SSRN https://doi.org/10.2139/ssrn.4865427

Jones, B., Jones, R., & Luger, E. (2022). AI 'Everywhere and Nowhere': Addressing the AI Intelligibility Problem in Public Service Journalism. *Digital Journalism*, *10*(10), 1731–1755. https://doi.org/10.1080/21670811.2022.2145328

Keller, P. *Generative AI and copyright: Convergence of opt-outs? - Kluwer Copyright Blog*. (2023). Retrieved 7 August 2024, from https://copyrightblog.kluweriplaw.com/2023/11/23/generative-ai-and-copyright-convergence-of-opt-outs/

Laranjeira de Pereira, J.R., *The EU AI Act and environmental protection: The case for a missed opportunity | Heinrich Böll Stiftung | Brussels office—European Union*. (n.d.). Retrieved 7 August 2024, from https://eu.boell.org/en/2024/04/08/eu-ai-act-missed-opportunity

Miller, G. (2024, February 20). *The Digital Services Act Is Fully In Effect, But Many Questions Remain | TechPolicy.Press*. Tech Policy Press. https://techpolicy.press/the-digital-services-act-in-full-effect-questions-remain

Nechushtai, E. *Could digital platforms capture the media through infrastructure? 2018*. Journalism, 19, 1043-1058 Retrieved 7 August 2024, from https://journals.sagepub.com/doi/abs/10.1177/1464884917725163

Nordic AI Journalism 2024, *AI transparency in journalism* https://www.nordicaijournalism.com/_files/ugd/efd447_06c5eb91a6084cf78da840af828ce845.pdf

Reporters Without Borders, 2024, *Countries signing Council of Europe's AI framework convention should not be tempted by self-regulation option* https://rsf.org/en/countries-signing-council-europe-s-ai-framework-convention-should-not-be-tempted-self-regulation

Siegmann Ch., Anderljung M., *The Brussels Effect and Artificial Intelligence: How EU Regulation Will Impact the Global AI Market TECHNICAL REPORT*, Centre for the Governance of AI, 78 (Aug. 2022).

Schjøtt Hansen, A., Bocyte, R., Krack, N., Dutkiewicz, *Blog series: AI regulation is overlooking the need for third-party transparency in the media sector* (2024) https://medium.com/ai-media-observatory/blog-series-ai-regulation-is-overlooking-the-need-for-third-party-transparency-in-the-media-sector-df843118c1fa

Tarkowski, A. *AI Act fails to set meaningful dataset transparency standards for open source AI – Open Future*. (2024). Retrieved 7 August 2024, from https://openfuture.eu/blog/ai-act-fails-to-set-meaningful-dataset-transparency-standards-for-open-source-ai/

Yordanova, K. *The Shifting Sands of Regulatory Sandboxes for AI* (2024) https://www.law.kuleuven.be/citip/blog/the-shifting-sands-of-regulatory-sandboxes-for-ai/

Zier, J. (2024, August 1). *"This Article is AI-Generated": AI Disclosure and Labeling for News Content*. Medium. https://generative-ai-newsroom.com/this-article-is-ai-generated-ai-disclosure-and-labeling-for-news-content-73d6e6cb6d50

7 Appendix: Updated AI4Media Strategic Research Agenda on AI for media (booklet)

In the following, we include as an appendix the booklet developed for the updated version of the AI4Media SRA. The booklet was released in July 2024 and is available on the project website.

AI4media
ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

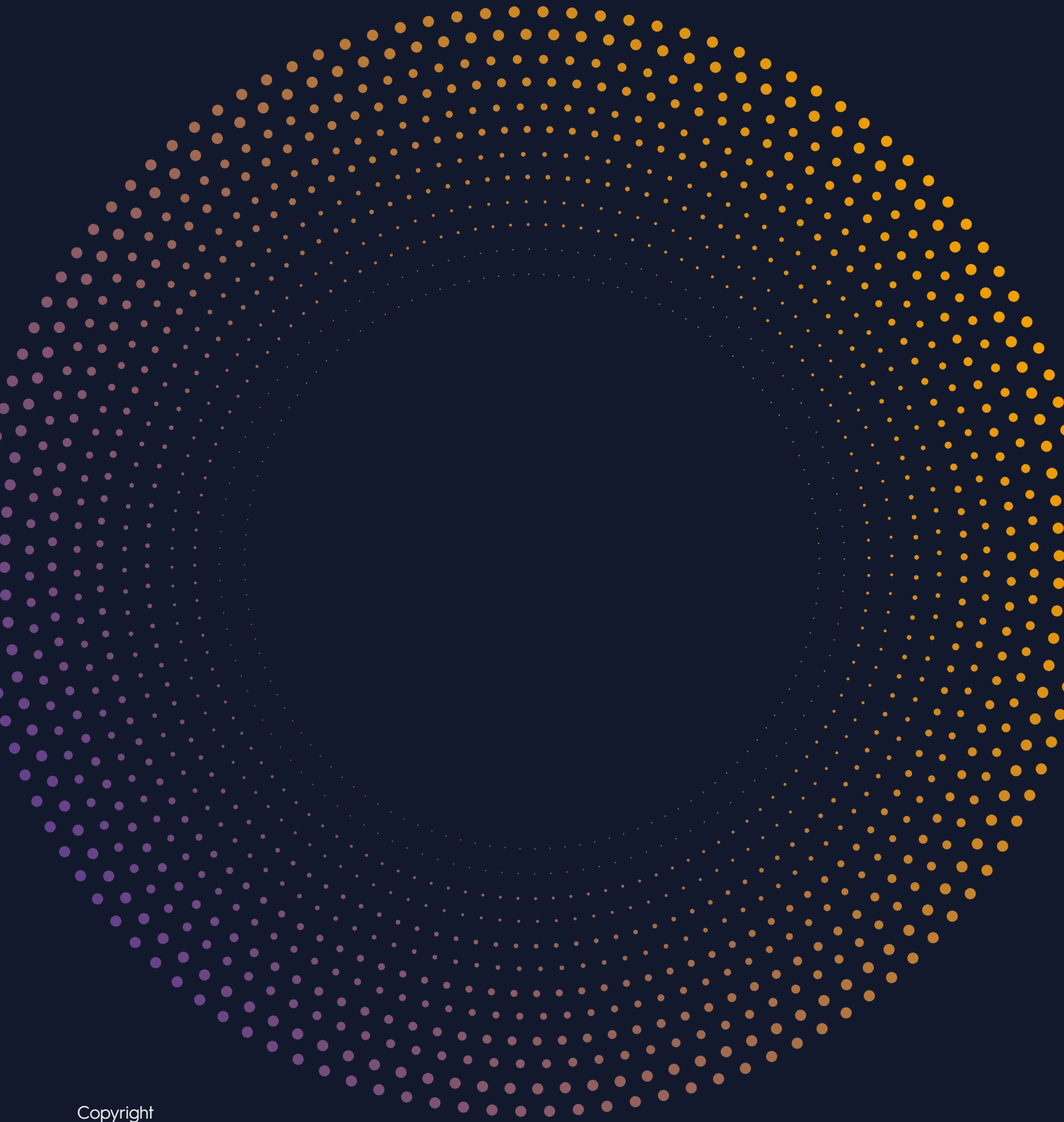
www.ai4media.eu
@ai4mediaproject

AI for the Media Industry
A Strategic Research Agenda from the AI4Media consortium

Updated edition - July 2024

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



Index of Contents

1. Foreword 06
2. Executive Summary 07
3. Introduction 09
4. Background: AI for the Media Sector 11
   - 4.1 Media AI landscape 11
   - 4.2 Overview of main opportunities & challenges of media AI 12
5. Context: A European Excellence Centre for Media, Society and Democracy 18
6. European Observatory for Media AI 22
   - 6.1 Context and need 22
   - 6.2 Research directions 23
   - 6.3 Expected outcomes 25
7. Main themes for core media AI research 26
   - 7.1 New Learning Paradigms and Distributed AI 27
     - 7.1.1 Context and need 27
     - 7.1.2 Areas of research interest 27
   - 7.2 Trustworthy AI 31
     - 7.2.1 Context and need 31
     - 7.2.2 Areas of research interest 31
   - 7.3 Content-centered AI 34
     - 7.3.1 Context and need 34
     - 7.3.2 Areas of research interest 34
   - 7.4 Human-centered and Society-centered AI 38
     - 7.4.1 Context and need 38
     - 7.4.2 Areas of research interest 38
   - 7.5 Expected outcomes 41
8. Generative AI and LLMs for the Media 42
   - 8.1 Context and need 42
   - 8.2 Research on legal, ethical and societal aspects 43
     - 8.2.1 Generative AI and the AI Act 43

8.2.2 Dataset diversity and quality 44
8.2.3 Challenging power dynamics and infrastructure capture 44
8.2.4 AI-generated content and copyright laws 45
8.2.5 Impact on epistemic welfare and the (dis)information ecosystem 46
8.2.6 Impact of Generative AI in media production and creativity 46
8.2.7 New displacement patterns and hidden labour 47
8.2.8 Good practices for the use of GenAI in the media sector 47
8.2.9 Generative AI Literacy for media professionals and media consumers 48
8.2.10 Environmental impact 49
8.3 Research on key technology aspects 49
8.3.1 Domain adaptation and fine-tuning of Large Models 49
8.3.2 Incorporating knowledge in foundation models 50
8.3.3 LLM hallucinations and LLM output quality 50
8.3.4 LLM unlearning 51
8.3.5 Embodied LLMs 51
8.3.6 Dataset biases and fairness 52
8.3.7 Enhancing the interpretability of Generative AI models 52
8.3.8 Explainable vision-language models 53
8.3.9 Resource-efficient vision-language models 53
8.3.10 Physics in vision-language models 54
8.3.11 Integration of new/multiple modalities in next-generation Large Models 55
8.3.12 Collection-wise Visual Question Answering 55
8.3.13 Generalizable evaluators for high-quality content generation 56
8.3.14 Face dataset anonymization 56
8.3.15 Enterprise context-aware Generative AI 57
8.4 Media and creative applications 58
8.4.1 Personalised LLMs for journalists 58
8.4.2 LLMs as an assistant research tool for media analysis 58
8.4.3 Indexing the entire World Wide Web with LLMs embeddings 59
8.4.4 Media agents 60
8.4.5 Audio-visual synthetic media detection 60
8.4.6 Content provenance markers 61

8.4.7 Ensuring transparency and the right to be informed through labelling the use of Generative AI in media production 61
8.4.8 Real-time music generation 62
8.4.9 Emotionally expressive avatars 62
8.4.10 Multimodal text instructable agents in interactive worlds 63

9. Use cases in media, society and politics 64
9.1 Introduction 64
9.2 AI4Media Use Cases 65
9.2.1 UC1: AI for Social Media and Against Disinformation 65
9.2.2 UC2: AI for News - The Smart News Assistant 66
9.2.3 UC3: AI in Vision - High quality Video Production and Content Automation 66
9.2.4 UC4: AI for Social Sciences and Humanities 67
9.2.5 UC5: AI for Games 68
9.2.6 UC6: AI for Human Co-creation 68
9.2.7 UC7: AI for Content Organisation and Content Moderation 69
9.3 Intersection of AI4Media Use Cases & AI Research Themes 69
9.4 Expected outcomes 70

10. AI education and AI skills development 72
10.1 International AI Doctoral Academy (AIDA) 72
10.2 AI4Media Junior Fellows Exchange Program 73
10.3 AI4Media open calls 73

11. Building the European Media AI Ecosystem 75

12. Conclusion 78

1. Foreword

Following a series of breakthroughs in the field of Artificial Intelligence, new technologies are emerging which are ushering a wave of revolutionary innovations in nearly all aspects of business and society; from transportation to finance, the fight against climate change, the media industry, journalism and politics. In all facets of economic and social life, AI is disrupting existing practices and creates significant opportunities for economic growth and societal prosperity while also involving considerable ethical challenges and risks.

Motivated by the opportunities, challenges, and risks that the wide use of AI brings to media, society and politics, AI4Media is building a wide Network of researchers across Europe and beyond, with a focus on delivering the next generation of core AI advances to serve the key sector of Media, making sure that the European values of ethical and trustworthy AI are embedded in future AI deployments, and reimagining AI as a beneficial enabling technology in the service of Society and Media.

This Strategic Research Agenda is a key contribution towards building this European Network of Excellence focusing on AI for the Media and Society. The objective is twofold: on one hand, it lays out the strategic plan for AI4Media's R&I activities, presenting the main research themes to be tackled by the consortium, explaining the current challenges, the research directions that need to be pursued to address them, the media industry applications, and the potential impact of this research. We hope this would be useful for AI researchers, media practitioners and policymakers. The second objective is to initiate and engage in discussions and debate with the broader community that could result in updated versions of this Agenda, thus making it a living document depicting the current status of AI research and applications for Society and Media.

A first version of the AI4Media Strategic Research Agenda was released in March 2023. An updated version is released in July 2024 aiming to reflect the significant progress made in the last year and a half with regard to Generative AI, fueled by the widespread adoption of Large Language Models and the rise of Large Multimodal Models. This new wave of AI innovation offers amazing opportunities for the media but at the same time raises significant concerns about its impact on creative industries, society, democracy, environment, etc. These aspects are explored in a new section dedicated to Generative AI and LLMs that examines legal, ethical and societal aspects, key technology trends, and relevant media and creative industry applications.

→ **Dr Yiannis Kompatsiaris**

AI4Media Coordinator

Director of the Information Technologies Institute, Centre for Research and Technology Hellas, Greece

→ **Dr Filareti Tsalakanidou**

AI4Media Project Manager

Research Associate in the Information Technologies Institute, Centre for Research and Technology Hellas, Greece

2. Executive Summary

AI4Media's vision is that of a **European Network of Excellence in Artificial Intelligence for the Media, Society and Democracy** that will glue together the pieces of the currently fragmented European AI landscape and promote a unique brand of European Media AI. AI4Media will build a network of experts, including both leading researchers in media AI from academia and research as well as top European media companies that use AI to enhance their operations and business opportunities. Together, they will address significant technical, legal, ethical and application challenges, aiming to address pressing needs of the media industry and significant societal problems.

The Media are already benefiting from AI advancements and AI-driven applications that can significantly facilitate, enhance or transform important tasks, including smart assistants, smart recommender systems, content personalisation, automatic content creation, multi-modal content search, multilingual translation, disinformation and manipulated content detection, social media analysis and trend detection, online debate analysis, forecasting and decision support-systems, and many more. Further **advances in AI have the potential to transform the media industry** and revolutionise how operations run and how content is created, delivered and consumed while they can also offer trustworthy solutions with a societal impact, aiming to improve political participation, increase social cohesion, equip citizens against disinformation, and encourage healthy debates and social interaction.

To realise this enormous potential of AI requires breakthroughs in several domains such as:

→ **Machine learning** (ML), aiming to address important challenges of current ML techniques, including learning with few data, learning on-the-fly, transfer of knowledge and optimal AI architectures. In addition, research should also focus on distributed AI systems running on heterogeneous devices but also disruptive technologies currently at the laboratory stage such as Quantum-assisted Reinforcement Learning.

→ **Content-centered AI technology**, valuable for the media industry and marketable as end-user services, such as multimedia metadata extraction, summarisation, and clustering, automatic audiovisual content generation and enhancement, linguistic analysis, and media-specific core technologies to improve learning performance.

→ **Human and society-centered AI technology**, to equip citizens and media professionals with a set of tools that can be used to counter the effects of media manipulation and disinformation, enhance the understanding of online debates, support the analysis of perceptions of social media and the effects of online data sharing, and improve local news understanding without being limited by language barriers.

→ **Trustworthy AI techniques**, that aim at providing a framework for the development of the technologies mentioned above that guarantees their suitability with respect to democratic and ethical values. Research should focus on issues of robustness against threats and malicious attacks, explainability of AI decisions, fairness and mitigation of bias of AI models, and techniques for privacy-preserving AI.

In addition to the aforementioned topics of interest, another research area where rapid progress has been observed in the last year and a half is that of **Generative AI (GenAI)**. The wide adoption of Large Language Models (LLMs) like ChatGPT and the rise of large multimodal models (LMMs) is transforming existing media industry workflows, essentially reimagining content creation and consumption. Research should focus on both technical challenges like domain adaptation, multimodality, explainability, AI hallucinations, resource-efficient models, dataset biases and fairness but also on ethical, legal and societal aspects like AI policy & regulation, copyright of AI inputs and outputs, power dynamics, labour displacement, impact on human creativity, disinformation, environmental impact, and more. At the same time, important applications of generative AI for the media industry should be studied like use of LLMs as personal assistants, interactive agents in virtual worlds, synthetic content detection, artistic creation, etc.

These research outcomes are integrated and evaluated in **real-world use cases**, aiming to address significant challenges currently faced by different media industry sectors and to highlight how AI applies throughout the media industry value chain, from research and content creation to production, distribution, consumption/interaction, performance and quality measurement. The use cases cover a variety of media and societal topics such as **disinformation, news research and production, organisation of media archives and content moderation, game design, human-machine artistic co-creation, and social science research.**

In parallel to delivering the next generation of AI research at the service of media, AI4Media aims to establish a **Media AI Observatory** to monitor the legal and technological landscape as well as the impact of media AI on the society, economy and democracy. The Observatory provides an overview of the existing EU policy and legal initiatives and their impact on future AI research for the media industry, analyses ethical, societal, environmental and economic concerns, and provides easy access to leading experts in this domain.

Implementing our vision of **AI as a human-centered, trusted, and beneficial enabling technology in the service of media and society**, requires supporting in practice the next generation of AI talent in Europe by offering opportunities for top AI education and skill development while also supporting entrepreneurship and innovative ideas. To this end, AI4Media established the **International AI Doctoral Academy**, a joint ICT-48 instrument to support world-level AI education and training for PhD/postdoc AI researchers. In addition, it provides **mobility opportunities** for young researchers and media professionals. And lastly, it funds and **supports SMEs, start-ups and research labs** that want to develop innovative applications and research for the Media. These activities will further strengthen the European AI research community.

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 and transform the industry. The AI4Media Network of Excellence aims to play an important role in this transformation, by bringing together leading research and industry players in this domain to strengthen the competitiveness and growth of the European media industry and increase Europe's innovation capacity in media AI. The second version of this Strategic Research Agenda crystallises AI4Media's research and innovation activities to materialise this vision.

3. Introduction

AI4Media aspires to become a European **Network of Excellence in Media Artificial Intelligence** that will:

→ Deliver the next generation of AI research and training at the service of media, society and democracy.

→ Make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments in the media sector.

→ Reimagine AI as a human-centered, trusted, and beneficial enabling technology in the service of Media and Society.

The ambition of AI4Media is to **put together the pieces of the currently fragmented European AI landscape** and establish the EU as a strong international competitor through a **unique European brand of Media AI**. To this end, the AI4Media network brings together Europe's leading researchers in AI media research, including both AI researchers that develop AI algorithms and tools for the media but also legal experts and social scientists whose research activities focus on the media sector and/or AI technology. In addition, top European media houses and content-related companies offer their unique perspective on industry challenges and needs, guiding research and innovation in the project.

This concentration of expertise makes AI4Media uniquely placed to drive forward AI science and its application in the media domain. Together, this community will build a new generation of AI technologies for the media sector that address real industry needs, are easily integrated in real-world media workflows, are aligned with European societal values and concerns, and adhere to European principles of Trustworthy AI. Reflecting this approach, this **Strategic Research Agenda** (SRA) provides a framework for coordinating AI4Media research and innovation activities.

The AI4Media Strategic Research Agenda comprises the following elements:

→ Development of an **AI for Media Observatory** to monitor the AI policy/regulatory landscape, AI societal impact, and AI technology trends, providing insights, analyses and recommendations for policymakers, industry and researchers.

→ **Research on four core AI areas** that will help reinforce and extend Europe's expertise in AI for Media. These include: new machine learning paradigms, trustworthy AI, content- centered AI, and human/society-centered AI.

→ **Research on Generative AI and LLMs** that will help the media industry to exploit the full potential and capabilities of this promising and rapidly adopted technology while successfully addressing current challenges, concerns and limitations (this is a new section included in the second version of the AI4Media Strategic Research Agenda).

→ **Real-world case studies** showcasing how to transform AI research into practical applications for the media industry with concrete impact to society and the economy.

→ **AI education and AI skills development** through the establishment of a prestigious European PhD programme on AI, a flexible mobility program, and open calls for funding research labs and SMEs working on media AI.

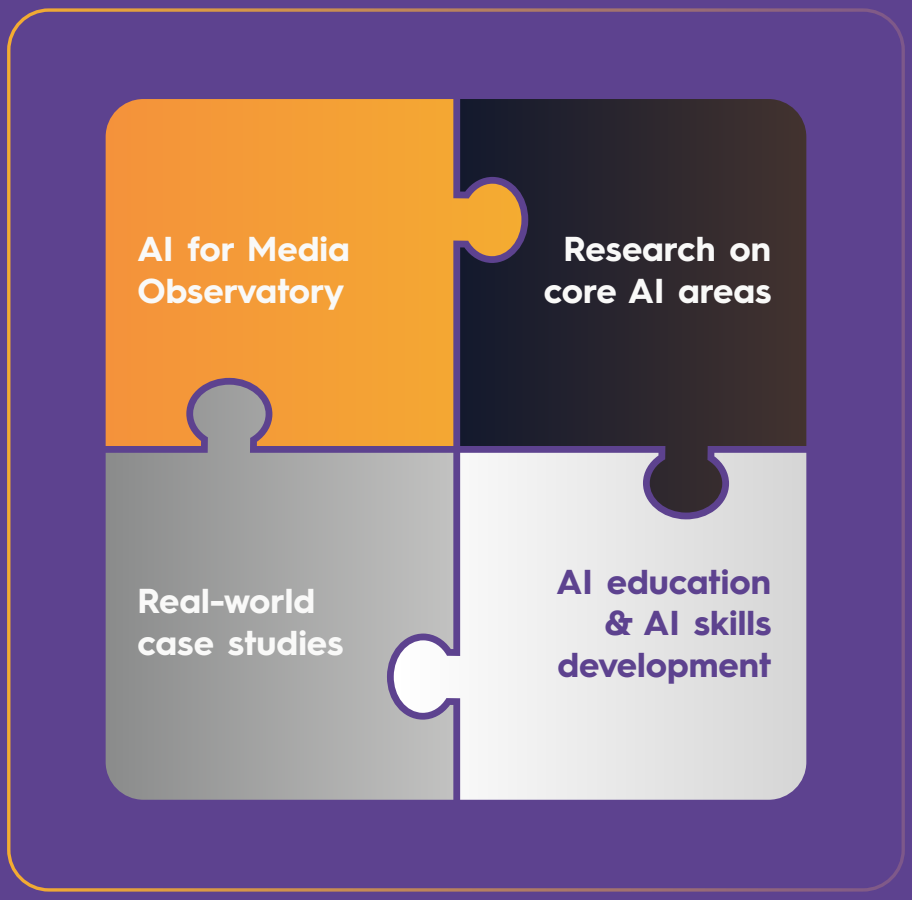


Figure 1

In addition, the document provides context for the SRA by providing an overview of the AI for Media landscape, discussing opportunities and challenges, while also providing more insights on the vision and ambition of AI4Media to become a European Network of Excellence on Media AI.

The AI4Media SRA was produced following a co-creation process that built on the AI4Media Description of Action and the **AI4Media Roadmap of AI for the media industry**, including a state of-the-art-analysis; a public survey addressed to AI researchers and media professionals across Europe and beyond; a consultation with representatives from the media industry, AI research community and policymakers; dedicated mini workshops, interviews and surveys with media professionals. The Strategic Research Agenda is a 'live' document, which evolves as the work progresses, responding to emerging industry and research needs and challenges and considering the expertise of the Network participants.

The second version of the Agenda includes a new section on Generative AI and LLMs that aims to summarise the technical, ethical, legal, societal and application challenges that have emerged after the latest wave of Generative AI breakthroughs.

4. Background: AI for the Media Sector

4.1 Media AI landscape

The world is changing. Following a series of breakthroughs in the field of Artificial Intelligence, new technologies are emerging which are ushering a wave of revolutionary innovations in nearly all aspects of business and society; from transportation to finance, the fight against climate change, the media industry, journalism, and politics.

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 automation[1], to provide solutions for and achieve breakthroughs in major world problems like poverty[2], climate crisis or cancer[3], to ensure equitable access for all[4], to increase productivity, innovation and creativity[5], to empower communities and strengthen democracy[6], 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, manipulation of public opinion through disinformation, violation of fundamental rights like the freedom of expression through questionable moderation choices, just to name a few) but also many others that we do not fully grasp yet[7]. 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[8].

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

2 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

3 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

4 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/

5 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/

6 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/

7 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

8 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

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 insights into 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.

AI will also transform the existing workflows of the media industry[9] by automating routine or tedious processes (from content creation and content search to automatic analysis of legal documents to ensure compliance with copyright standards); 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); improving audience analysis and user profiling to offer better content and services to users; 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 while it can help legacy media to be more competitive in a digital marketplace that is currently dominated by large platforms.

Empowered by AI-based support across the value chain, the media have the power more than ever before to nurture a democratic society, to improve our lives and support our creativity. However, at the same time, they can lead to polarisation and crisis, amplified by the use of the same AI. For example, there is increasing concern that the power of social media platforms combined with the large-scale automation capabilities offered by AI could prove to be detrimental to individuals, society and democracy. These concerns are fueled by the events of the last few years, where disinformation campaigns were found to have played an important role in shaping public opinion and the results of elections around the world.

4.2 Overview of main opportunities & challenges of media AI

As mentioned above, the media & entertainment industry (news, film/TV, music, games, social media, advertisement, publishing etc.) is already benefiting 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.

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

In the following, we briefly discuss the main opportunities for the use of AI in the media sector but also relevant challenges. This summary is based on the state-of-the-art analysis conducted as part of deliverable **D2.3** "AI technologies and applications in media: State of Play, Foresight, and Research Directions", which reviewed and analysed in depth 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. This analysis 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. At the same time, the use of AI in the media comes with significant challenges and risks that should be efficiently addressed to enjoy the benefits of this technology.

Opportunities for the use of AI in the media sector

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. The transformative role of AI has already started to manifest in many of these areas, with important breakthroughs over the last few years.

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, 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. To win the race, media companies invest large amounts 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).

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, creation of art 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, complex voiced questions or search based on images & sketches. 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.

Multilingual NLP. Natural language processing 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. Multilingual 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.

Audience analysis. Understanding what the audience wants or needs or how the audience feels is a top priority. 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.

Challenges for the use of AI in the media sector

The previous analysis reveals the vast potential of AI to bring positive change to the media industry sectors. However, with high potential also come significant challenges and risks.

AI explainability. Currently, AI systems are mostly black boxes, without being able to explain why they predicted the success of a film or verified the authenticity of a video. 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.

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 large amounts of Internet data that are produced in the biggest countries, in languages with higher linguistic footprint, and by communities with large representation, or mainly by men[10], 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[11,12]. To address similar problems, new techniques have been proposed aiming to enhance fairness of AI models.

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.

High-quality data for AI training. Large volumes of real and high-quality data are required for training AI models for the media industry. AI researchers are faced with significant challenges when it comes to this issue: low dataset diversity, lack of regionally relevant data (smaller languages/ dialects, locally relevant named entities, etc.), complex data gathering processes, closed datasets, GDPR concerns, dataset bias, and proper benchmarking.

10 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.

11 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/.

12 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.

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 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. Many media companies still 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, pursue useful collaborations and fully exploit AI's potential for the media.

AI applications for different media industry sectors

The applications of AI for the media industry are numerous. Figure 2 attempts to summarise the relevant applications per industry sector, focusing on news, film/TV/streaming, music, games, social media, advertising, and publishing. As can be 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 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 mentoring in fanfic communities
- Copyright management
- Forecasting (sales, trends, content appeal, etc.)

Figure 2: A summary of AI applications for the media and entertainment industry.

5. Context: A European Excellence Centre for Media, Society and Democracy

Motivated by the challenges, risks and opportunities that the wide use of AI brings to media, society and politics, AI4Media aspires to become a **Centre of AI excellence and a wide Network of researchers across Europe and beyond**, that will:

→ Deliver the next generation of AI Research and Training at the service of media, society and democracy.

→ Make sure that the European values of ethical and trustworthy AI are embedded in future AI deployments in the media sector.

→ Reimagine AI as a human-centered, trusted, and beneficial enabling technology in the service of Media and Society.

The ambition of AI4Media is to connect the fragmented European AI landscape and establish the EU as a strong international competitor through a unique European brand of Media AI. To achieve its overall ambition, AI4Media will be based upon 6 main pillars, as shown in Figure 3.

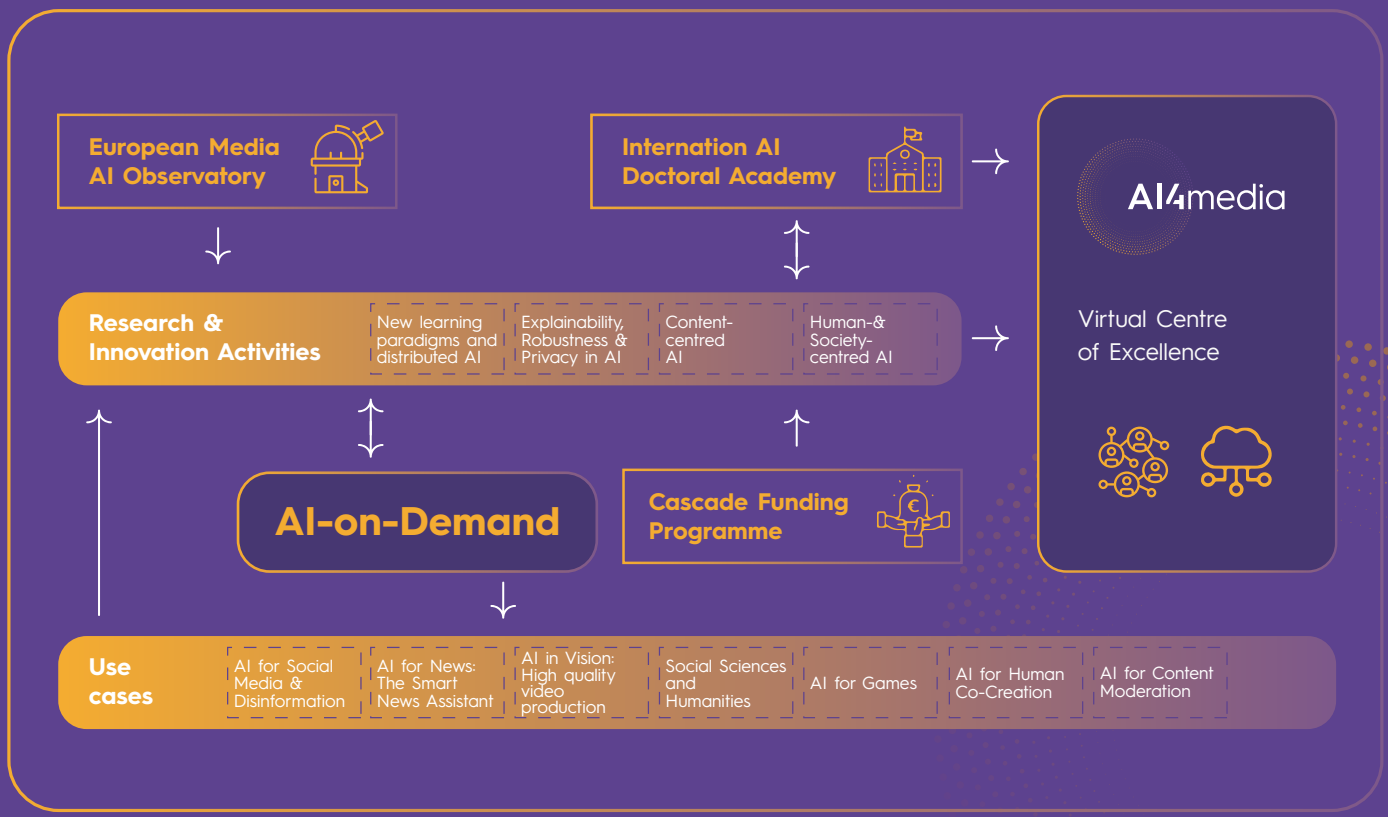


Figure 3: The AI4Media pillars

→ A **European Media AI Observatory** that will set a research and innovation agenda for media AI, monitoring the disruptive potential of emerging technologies for the society and economy and offering relevant policy recommendations.

→ An **intensive research and innovation activity plan** in four core areas of Media AI where Europe can acquire a competitive advantage: (1) emerging machine learning paradigms and AI at the Edge; (2) AI fairness, explainability, robustness, and privacy; (3) AI for multimedia content analysis and creation; and (4) Human-centered and Society-centered AI technologies. In addition to these areas, recently the focus also shifted on the potential and opportunities brought by Generative AI and LLMs. The technologies to be produced as a result of this research will enrich the AI-on-Demand platform.

→ A **portfolio of seven use-cases**, aiming to directly apply the technologies developed in the project to strengthen the European Media sector and society. The use cases that will be implemented through the close collaboration of AI researchers, media companies and user groups of media professionals aim to bring AI expertise to real-world applications of the media industry, by exploiting mature technologies developed as a result of the core AI research performed within the project. Covering a variety of topics such as disinformation, news research and production, media moderation, organisation of audiovisual archives, game design, human-machine co- creation, and social science

→ A targeted **programme of cascade funding** to increase engagement of external actors in AI research and applications and build an ecosystem around the network, in turn benefiting from it and bringing innovation to the market. The programme will include two open calls, funding in total 20 research or application projects with up to 50K euros per project.

→ An **International AI Doctoral Academy** that will nurture a new generation of AI talent by offering educational courses, student exchanges, AI mellontology symposia, summer schools, etc. Building upon collaborations between prominent academic institutions and the industry and cooperating with other networks of excellence to capitalise on aggregated expertise and resources, AIDA aims to achieve European academic excellence and industry relevance, attract young talent and provide incentives for it to stay in Europe.

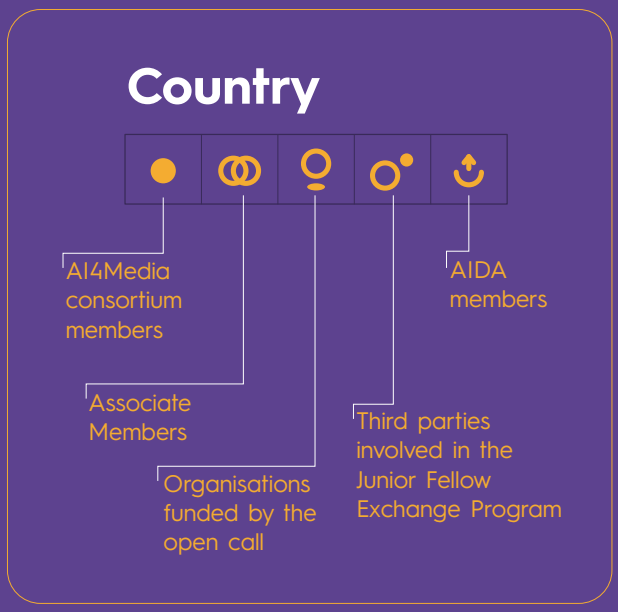
→ The **AI4Media Virtual Center of Excellence**, which will function as a portal for all Media AI research and innovation activities in Europe, aiming to glue together the pieces of the fragmented European AI landscape.

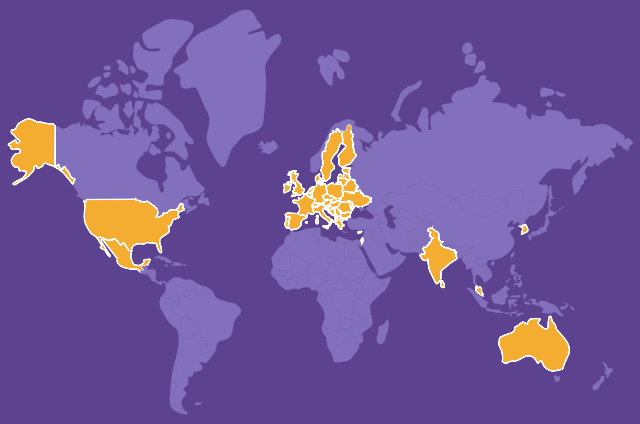
→ The **AI resources** developed in the project will be published in the AI Catalogue of the **AI-on-Demand platform** while selected AI components will be integrated in the **AI4EU Experiments** marketplace.

The AI4Media network

The AI4Media vision is to become a European Powerhouse of Media AI. Motivated by a shared desire to establish European leadership in the development and deployment of AI for the media sector, AI4Media has established a network of excellence that connects leading researchers, ICT industry and top media organisations across Europe.

The AI4Media network currently consists of **30 partners** (9 universities, 9 research centres, 12 industrial partners) and more than **100 associate members**, aiming to bring together the currently fragmented European AI landscape in the field of media, and foster deeper and long-running interactions between academia and industry. The network also includes research organisations and SMEs either funded by the project's open calls or participating in research exchanges as part of the AI4Media Junior Fellows Exchange Program.





Country	AI4Media consortium members	Associate Members	Organisations funded by the open call	Third parties involved in the Junior Fellow Exchange Program	AIDA members
Australia		2			
Austria	1	8		2	2
Belgium	2	4		3	2
Bulgaria	1	2	1		1
Cyprus		2			1
Czech Republic		1		1	2
Denmark	1	3	1	1	2
Estonia					1
Finland					1
France	4	11		5	6
Germany	3	7	3	6	8
Greece	3	9	3	3	4
Hungary					3
India		1			
Ireland	2	2		2	2
Israel			1		1
Italy	4	8	1	6	11
Lithuania					1
Luxembourg					2
Malta	1			1	1
Mexico				1	
Netherlands	2	4		4	3
North Macedonia				1	
Norway		2			
Poland		1			2
Portugal	1	1	1	2	1
Rep. of Korea		1			
Romania	1	12	2	2	1
Serbia		2		2	
Singapore		1			
Slovakia		1			2
Slovenia		1	1		
Spain	1	11	3	9	9
Sweden		1			3
Switzerland	2	6		5	3
UK	1	3	3	7	2
USA		4		1	

Unique selling point

AI4Media brings together a critical mass of top AI researchers, media professionals, social scientists and legal experts to create a Network of Excellence and a European Powerhouse in Media AI that will deliver the next generation of AI technologies for the Media Industry and will reimagine AI as a human-centered, trusted and beneficial enabling technology that can be used to offer innovative solutions to major challenges facing the media, the society and democracy.

Next generation AI for the Media

Core research

Multi-modal content

Human and Society in the centre

New generation of machine learning systems

Trustworthy AI solutions

Real-world applications

Digital content verification

Automated game design

News production automation

Audiovisual archives exploration

Content moderation

Human-machine artistic co-creation

Impact of policy & regulations

Monitoring of EU regulatory landscape

New policy recommendations

Societal concerns

Analysis of media AI impact & societal concerns

Education, training and outreach

AIDA ARTIFICIAL INTELLIGENCE DOCTORAL ACADEMY

Junior Fellows Exchange Program

AI4media OPEN CALLS

6. European Observatory for Media AI

6.1 Context and need

The world needs to be able to make sense of AI developments and their impact on individuals, society, and economy. The fragmented European AI landscape is in even greater need of an institution that will monitor, aggregate, study, and interpret relevant information.

To this end, AI4Media combines expert technical knowledge, industry experience, socio-economic analysis, and legal analysis to monitor technological and industrial developments of media AI, assess their impact on society and the economy, and align with policy recommendations, self-regulatory initiatives and ethical considerations.

Research in this direction includes:

→ Analysis of regulatory and policy landscape in Europe and beyond.

→ Analysis of AI for media technology trends and exploration of relevant opportunities, challenges and risks for the industry.

→ Analysis of the societal, political, and economic impact of media AI technology and applications.

→ Development of policy recommendations to address the challenges encountered by media companies and AI researchers.

This research is conducted in the context of Work Package 2.

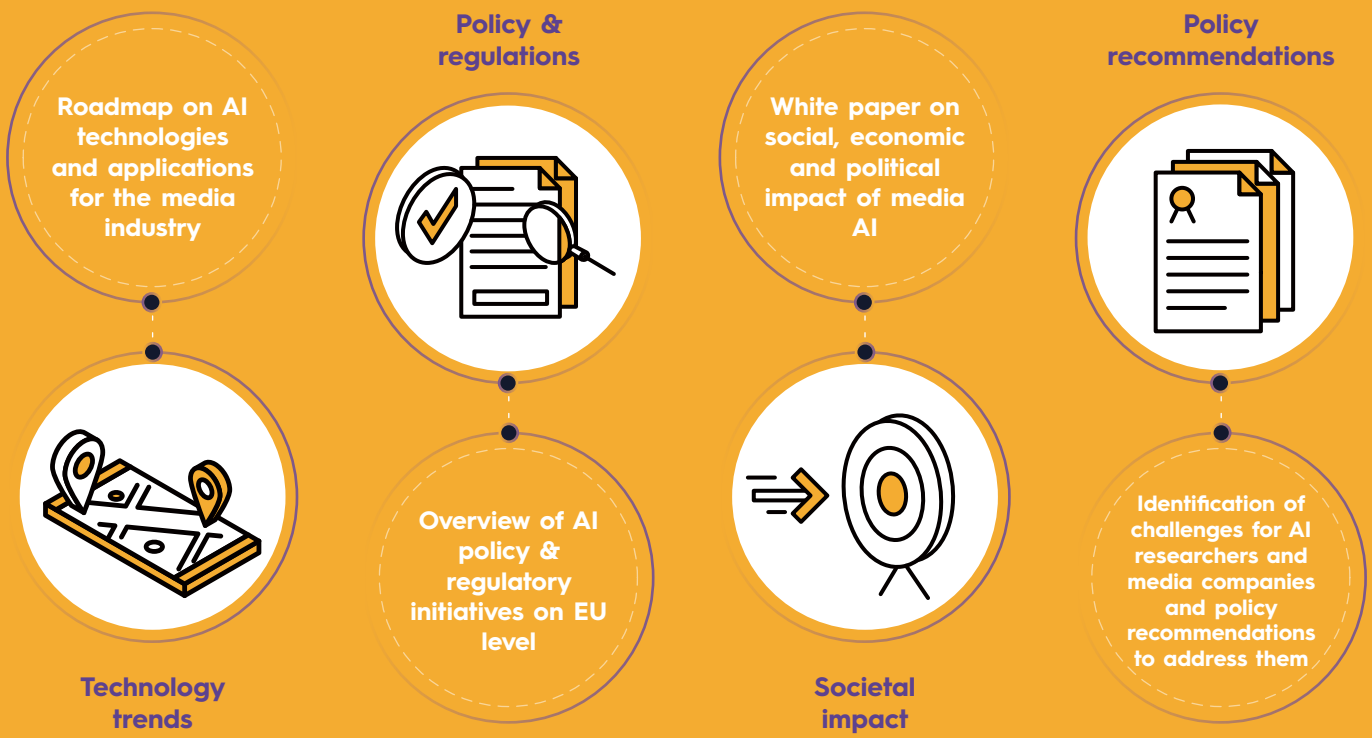


Figure 4: Media AI Observatory - main research directions

These research efforts culminated in the establishment of the **European AI Media Observatory**. Building upon the joint effort during AI4Media and using the project as a jumpstart, the Observatory will remain after the project completion as a long-term point of contact for European actors. Through the Observatory, media, research, and policy organisations will be able to acquire policy, technology, and application analyses and recommendations. Beyond the 4-year research agenda set by AI4Media, the Observatory will serve to maintain the gains established during AI4Media in the long run.

6.2 Research directions

Monitoring of EU and international AI policy and regulatory initiatives with a focus on the media sector

→ Challenge

The fragmented nature of EU and international policy and regulatory initiatives on AI triggers a hard-to-understand framework that requires analysis and explanation, especially concerning their impact on research activities that involve the development of tools and applications for the media. In addition, forthcoming EU policy and regulatory initiatives could potentially have a profound impact both on media-related AI research but also media companies' commercial and non-commercial activities.

→ Research

The focus is on monitoring and mapping the landscape of EU and international AI policies concerning AI ethics initiatives, including trustworthy AI, Intellectual Property rights, Safety and Liability regimes, use of AI in education, culture, and audio-visual sector, the AI Package including the AI Act proposal, the Digital Services Act Package, and the Data Package. Complementing this comprehensive mapping, the potential impact of the anticipated EU regulatory initiatives on the field of AI for the media will be also explored and analysed.

→ Expected impact

The mapping and ongoing monitoring of the AI policy and regulatory landscape allows AI researchers and the media industry to understand which legal framework affects them and in what way. Additionally, the analysis of potential gaps in these frameworks can be utilised by EU policymakers to affect and improve both legislative proposals that are in the process of adaptation at the moment but also forthcoming initiatives.

Roadmap on AI technologies & applications for the media industry

→ Challenge

While the media industry is already benefiting from AI advancements that can transform important tasks across the media value chain, the adoption of AI brings significant challenges and risks for the industry and society. Mapping the highly complex and evolving landscape of AI technologies for the media is a challenging task that has only been partially addressed, by focusing on specific sectors or aspects, thus failing to accurately grasp the whole picture.

→ Research

For the analysis of the AI for Media landscape, a multi-dimensional and multi-disciplinary approach is required, involving European AI researchers and media professionals. Three main tools are used: (i) a multi-disciplinary state-of-the-art analysis of relevant roadmaps, surveys, review papers and opinion articles performed by AI experts, social scientists, ethics and legal experts and media industry practitioners; (ii) a public survey targeted at AI researchers and media professionals; and (iii) development of a series of white papers on the future of AI for the media, focusing on different AI technologies and applications and on different media sectors.

→ Expected impact

By offering an in-depth look on media AI, the aim is to help i) the media industry discover the opportunities offered by AI to transform media workflows, assist media professionals and enhance the user experience in different sectors; ii) AI researchers to understand industry needs and recognise emerging research trends; and iii) social scientists and policymakers to comprehend facilitators, challenges and risks of media AI, resulting in more effective policies.

Analysis of critical societal concerns of media AI impact

→ Challenge

The media industry is rapidly adopting AI solutions across different parts of the media cycle (e.g., producing content, curating and distributing content, or moderating the engagement with media content). While several existing reports have focused on the implications of AI more generally or on specific applications or sectors, there is a need to gain a more informed and detailed overview of the potentials and challenges faced by the media sector as a whole when adopting AI, aiming to better understand the impact of media AI.

→ Research

To study and discuss the impact of media AI, two steps are envisioned: first, a systematic literature study of papers and reports from both industry and think tanks as well as academic journals, aiming to provide an overview of the state of discussion of AI for media, including the specific concerns raised in these discussions (technical, political, social, or economic). Second, a series of workshops with media industry representatives that focus on specific challenges identified in the literature review, aiming to qualify and expand on the initial literature based insights.

This research will lead to a whitepaper analysing the impact of media AI in society and the media sector specifically, identifying challenges and providing recommendations.

→ Expected impact

The literature study provides a 'reader's guide' for media professionals, AI developers working in the media sector, and researchers interested in AI and media. Identifying concrete challenges and providing recommendations for each challenge, it can help inform policy work, help professionals navigate in the AI landscape, and produce awareness of these challenges and the potential societal impacts of AI in media.

The industry workshops provide an opportunity for the media sector to directly engage with these challenges and in a more fine-grained manner identify where and how these challenges affect them and what responsible solutions to these challenges might be proposed. They, as a result, become active participants in shaping the agenda for responsible AI in the future.

Delivery of policy recommendations for media AI

→ Challenge

The current European legal landscape lacks the specific provisions, regimes, and policy framework concerning the use of AI applications and tools in the media sector. This absence of guidance creates important challenges for accomplishing the EU's vision on Trustworthy AI.

→ Research

To provide comprehensive pilot policy recommendations to meet the demand for clarification on AI use in media, we first need to identify the challenges for the use of AI applications in the media sector: challenges for media companies, for academia and researchers but also legal and societal challenges. The identification of challenges is based on the work performed under the previous research themes: overview of EU policy/legislative landscape, overview of landscape of AI technologies for the media, public survey on media AI, and overview of media AI impact.

The next step is to formulate a set of useful pilot policy recommendations to address the identified challenges, following a co-creation process that involves consultation with experts from different backgrounds: technology experts, social scientists, legal experts, media industry stakeholders, civil society representatives, policymakers.

→ Expected impact

This research will help raise awareness in the media sector about the importance of AI ethics. It will also help policymakers understand the challenges within

the media industry and academia and act accordingly while drafting legislation, policy frameworks, and regulations. While focusing on Europe, there is potential to also inspire other legal regimes in an affirmative way.

Establishment of the AI Media Observatory

→ Challenge

Currently, the information (case studies, journal articles, etc.) about AI in the media sector is scattered across multiple sites and platforms, which makes it difficult for media professionals, researchers and policymakers to efficiently navigate this information landscape. Equally, the growing amount of information produces the need to offer overviews of what challenges are the most urgent at this time.

→ Research

AI4Media will establish an AI Media Observatory aiming to monitor, aggregate, study, and interpret relevant information on media AI related topics, with the purpose to support a better understanding of media AI developments and their impact on society, economy, and people. The observatory enables this in three ways.

First, by offering an information one-stop-shop, where relevant content relating to AI for media will be curated, categorised and made searchable according to both genres (e.g., blogs, reports, news, forecasts, surveys, etc.) and on topics (e.g., trends, disinformation, policy etc.). Second, by providing an 'expert directory', where relevant experts in the field can be featured and contacted by, for example, the media industry or civil society. This will help stakeholders navigate this landscape of not only information but also experts. Third, by producing online filmed interviews or roundtable discussions with relevant experts, which can help provide easy overviews of what the core challenges of AI in media is at the moment and enable access to expert opinions in an easily digestible format.

→ Expected impact

The Observatory will positively impact the access to relevant knowledge on AI for media for industry, policymakers, researchers and civil society, by collecting it on one platform and also providing easy access to contact relevant experts in the field.

6.3 Expected outcomes

- **Overview & analysis of the AI policy initiatives on EU level**
- **Roadmap on AI technologies and applications for the Media Industry**
- **Online survey on AI for the Media Industry**
- **White paper on the social, economic, and political impact of media AI technologies (part 1, part 2)**
- **Pilot policy recommendations for the use of AI in the media sector**
- **AI Media Observatory**

Unique selling point

The AI for Media Observatory will become a reference point for monitoring EU policies and regulations, technology trends, and impact of media AI. It will also provide insights, analyses, and recommendations to policymakers, media industry, and research organisations from a multi-disciplinary panel of experts.

7. Main themes for core media AI research

To deliver the **next generation of core AI research for the media industry and society** and reinforce and extend Europe's expertise in media AI, the AI4Media consortium focuses its efforts in four crucial areas of core AI research:

- New machine learning paradigms and AI at the Edge;
- Trustworthy AI, exploring AI robustness, explainability, fairness and privacy;
- Content-centered AI, focusing on multimedia content analysis and creation;
- Human-centered and Society-centered AI technologies, focusing on the analysis of online debate, information production and consumption, and social media experience.

These research themes are explored in the context of Work Packages 3, 4, 5, and 6, respectively. Each research theme comprises several research directions, aiming to keep up with existing and emerging AI research trends but also address the needs and challenges of the media industry (Figure 5).

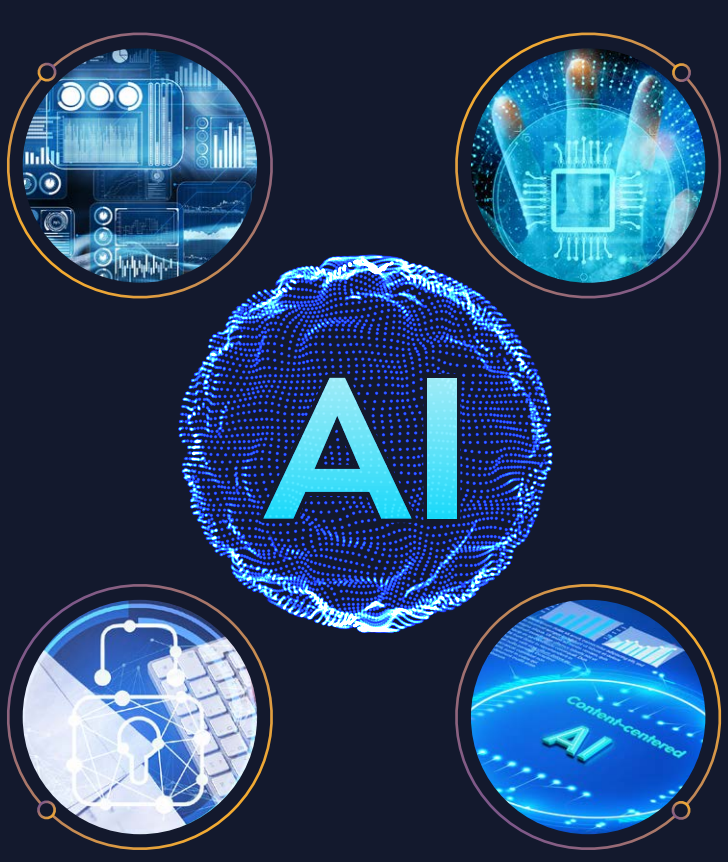


Figure 5: Core AI research for the media - Main research themes and research directions.

7.1 New Learning Paradigms and Distributed AI

7.1.1 Context and need

In the context of the current AI revolution, several core AI technologies that challenge the existing learning paradigms need to be explored. Models that adapt and learn on-the-fly, families of models that "teach" each other, algorithms that are able to learn from very few data and algorithms that are able to replace humans in selecting the optimal AI architecture to train for a given task, are all research propositions that have already begun to give promising results and may alter the way AI works in the very near future.

Distributed AI systems working on heterogeneous devices and powered by advanced communications like 5G challenge the paradigm of cloud computing which aggregates user data, centralises computation, and feeds the results back to user devices. AI at the Edge is often hailed as the next major evolution of AI, opening a window of opportunity for Europe to acquire a competitive advantage. Disruptive technologies currently at the laboratory stage, such as Quantum-assisted Reinforcement Learning, should also be explored as they are emerging.

In the following, we summarise the main areas of research interest under the New Learning Paradigms research theme. The aim is to efficiently address the current limitations of learning approaches and improve learning performance and speed.

7.1.2 Areas of research interest

Lifelong and on-line learning

→ Research challenges

Lifelong/on-line learning enables deep neural networks to learn from streams of data, while storing only a limited footprint of past data. The main challenges are to: (i) make the data streams usable for the training of deep learning models, (ii) maintain a good performance level for both past and new data, and (iii) update deep models without full retraining.

→ Research directions

The first challenge can be addressed by proposing novelty detection methods, which make use of past knowledge to label new data, notably by incorporating contrastive learning techniques to differentiate between classes. The second challenge can be tackled by reducing the bias towards new data, which occurs due to imbalance in their favour. The proposed techniques focus on past data in order to boost their deep representations, and thus make the lifelong learning process both stable and plastic. The third challenge can be addressed by reusing knowledge encoded in past models as much as possible in order to reduce the complexity of the lifelong/on-line training process.

→ Expected impact

The knowledge relevant to the media sector evolves fast and models used to extract it should be able to follow this pace. Lifelong/on-line learning techniques can be used to: (i) make sense of large and unlabelled streams of data, (ii) classify new entities in real-time via a fast update of the underlying deep models, and (iii) identify potentially problematic content via out-of-distribution detection.

Manifold learning and disentangled feature representation

→ Research challenges

Learning features manifolds of unstructured data (e.g., images) where semantic concepts lie is paramount for systems that recognize and synthesise new data (e.g., for helping media practitioners) but hard without manual labelling and supervision. Unsupervised manifold learning using rich multimodal pre-trained models aims at raising these limitations, and allowing for discovery of meaningful and useful factors without subjective and laborious human labelling.

→ Research directions

These challenges can be tackled along two research lines. First, by optimising in an unsupervised manner (i.e., without the need of laborious human labelling) controllable generative paths allowing for intuitively

controlling the generation of media content. Moreover, by incorporating recent pre-trained Vision-Language models (such as CLIP) that can serve as a cheap and rich source of supervisory signals towards the discovery of controllable generative paths, allowing generation controllable by guidelines given in natural language.

→ Expected impact

Research in this area allows for generative methods (e.g., of facial imagery) that are more controllable by taking advantage of vast pre-trained Vision-Language models. This may help in creating more robust and democratised models/datasets by enhancing/augmenting datasets with controllably generated samples of under-represented classes.

Transfer learning

→ Research challenges

Training a neural network usually requires very large computational and energy costs, which can be un-assumable for some media practitioners. Transfer learning research aims at lowering these costs by studying the reuse of already trained AI models for new tasks in new domains, thus avoiding training neural networks from scratch when possible.

→ Research directions

These challenges can be tackled through two research lines: (i) methodology studies, and (ii) creation of robust, reusable models. Methodology studies compare the AI models' performance, computation time, and/or power and energy consumption in diverse domains and tasks, when applying different transfer learning methods. The second research line is focused on training AI models that are as easily reusable and adaptable as possible, by learning domain-invariant knowledge or using domain alignment techniques when reusing models.

→ Expected impact

Transfer learning methods allow more media outlets and professionals to easily and successfully reuse and adapt existing AI models for their specific needs, such as face and person detection, image labelling and classification, information retrieval or image and video manipulation tasks.

Neural Architecture Search

→ Research challenges

Neural Architecture Search (NAS) has been successfully used to automate the design of deep neural network architectures; however, why these architectures work well, how similar are the architectures derived from different search strategies, how to design and access the search space in an efficient and unsupervised way, and how to fairly evaluate different auto-designed architectures remain open research challenges.

→ Research directions

Parameter sharing based OneshotNAS approaches can significantly reduce the training cost of AI models. However, there are still two issues that need to be solved in the development of lightweight NAS: i) the performance of the network sampled from the supernet is inconsistent with the performance of the same network trained independently. The solution is to investigate a better evaluation strategy and an improved ranking of candidate performance; ii) the existing performance prediction benchmarks do not yet focus on the evaluation of networks from the perspective of different search spaces. The way to go is to propose solutions which are learnt for practical real-world applications.

→ Expected impact

Investigating NAS in the "real-world" has received limited attention. Our research will allow searching for an optimal deep neural network architecture on an entirely novel task or dataset, facilitating the deployment of such models in the media sector. We will establish a set of best practices and look into optimal architectural patterns, augmentation policies, and hyperparameter selection.

AI at the Edge and distributed learning

→ Research challenges

AI at the edge makes it possible for AI applications to run at end devices (e.g., smartphones) aiming to protect user privacy and scaling computations to the network. Challenges are due to i) user device limitations, e.g., limited computing power and storage, ii) network limitations such as delay and unreliable intermittent communication, iii) limitations of the distributed learning paradigm such as limited data per device and convergence issues during training, and iv) security issues that arise both from participating users (e.g., model/data poisoning attacks) and coordinating nodes (e.g., reconstruction attacks).

→ Research directions

An important line of research is the adaptation of AI models to the limitations of end user devices, which is crucial for the deployment of AI at the edge. This approach includes (a) compression of trained models to fit on mobile devices, (b) efficient learning and inference from limited data, such as by focusing only on the data samples that are most useful to learning subject to resource constraints, as well as (c) support of lightweight AI paradigms under limited resources, for instance to train within edge devices.

At the same time, it is important to focus on collaborative learning among devices, which safeguards the privacy of personal data and enables scaling computations to the whole network. Approaches that both rely on coordinating nodes (federated learning) and are completely decentralised are relevant. It is also important to develop the necessary software tools and middleware to train and deploy AI models directly at user devices.

→ Expected impact

AI-at-the-edge will enable AI applications at mobile devices that are highly useful to media professionals while foregoing centralised services. This is key to protecting journalists and their work, as it lets them maintain privacy and control over their personal data, especially if they are working under authoritative regimes. Potential applications involve processing multiple types of data, ranging from text for fact-checking, sound for synthetic speech detection, and image / video for deepfake media detection.

Deep quality diversity

→ Research challenges

Evolutionary computation has been particularly powerful in numerical optimisation tasks, where a clear optimal solution exists. However, in subjective domains such as media and games, evolution driven by objectives alone is not sufficient. Quality-diversity (QD) algorithms are a trending new solution to this but the challenge of defining appropriate dimensions for both quality and diversity is almost as challenging as the search task itself.

→ Research directions

The problem of defining appropriate dimensions for quality and diversity in QD algorithms and searching efficiently along these dimensions is expected to be addressable via deep neural network architectures. On the one hand, deep networks can transform exploration, aiming for both diverse and high quality outcomes. The representation learning power of deep networks can be used as a way to define diversity along latent vectors provided by trained models.

On the other hand, QD algorithms can benefit optimisation of neural network weights (and possibly architecture) as well by promoting networks that output diverse results along one or more subjective dimensions. Moreover, normal backpropagation towards minimizing error can be coupled in sequence or in parallel with QD search to ensure that QD does not over-explore.

→ Expected impact

By coupling deep learning and QD evolutionary search, both of which are particularly popular solutions in their respective subfields of computational intelligence, we expect important breakthroughs in both subfields. Specifically, these approaches are expected to lead to novel media pipelines that produce unexpected but valuable artistic output such as visual art constrained by a semantic prompt but driven towards visual diversity (e.g. colors present).

Learning to count

→ Research challenges

When the "unlabelled" (i.e., new) data are different in nature from the training data, the trained classifiers are not accurate at estimating class proportions in new data. Research on learning to quantify (LtQ) aims at developing methods for estimating class proportions in new data even when these new data substantially differ from the training data.

→ Research directions

These challenges can be tackled in two main ways, i.e., (i) developing new LtQ methods, and (ii) developing robust experimental protocols for testing these methods on data. New LtQ methods can be built by studying the different ways in which unlabelled data may be different from the training data, and by building different LtQ methods each tailored to these different ways. The new experimental protocols to be defined must be able to generate test data that simulate the different ways in which unlabelled data may be different from the training data.

→ Expected impact

LtQ methods allow one to train estimators of the class proportions in new data. This is useful in all application contexts in which we are interested in aggregate data (rather than in individual data), such as the social sciences or political science; in these cases, predicting to which class a data item belongs is less interesting than estimating the fraction of data items that belong to a certain class. Additionally, prediction at the aggregate level is often interesting when privacy preservation is important, since it allows inferring interesting knowledge at the aggregate level while preventing undesired inference of knowledge about individuals.

Quantum-assisted reinforcement learning

→ Research challenges

Reinforcement Learning (RL) and other AI methods are computationally demanding tasks. Their development in recent years has been motivated by powerful hardware advances that have created new hardware systems custom fit to execute these methods. Following this line, having access to novel technologies of computations will allow the execution of more powerful algorithms and processing of larger sets of data, which is currently limiting their development.

→ Research directions

Quantum computation is currently being explored as a potential candidate to provide high performance capabilities in the next few years. This task requires the study of its performance on given datasets, but also the development of novel methods and algorithms.

We will explore Quantum algorithms used in the training of a RL model and the potential interaction of conventional computers with Quantum computers in solving hard problems such as model training. We will provide new hybrid algorithms and strategies tested in advanced computational systems – including real Quantum Hardware – that may elucidate the role that Quantum computation will play for AI applications.

→ Expected impact

Current ML systems are powerful combinations of hardware and algorithms. This is the result of a long process of development based on conventional computational processors. As a technology with a potentially large impact on any computationally hard problem, Quantum computation has to follow a similar track. Current efforts in Quantum ML are identifying the limitations of classical ML methods on which Quantum technologies can provide an advantage. Candidates include large scale problems of either classical or Quantum datasets, with applications to the media (e.g. multispectral image processing).

Unique selling point

AI4Media will provide the AI community with a set of novel open source tools covering new learning paradigms, including learning at the edge and quantum assisted learning, that can be easily integrated within existing AI pipelines and can be used by the community for implementing their use cases.

7.2 Trustworthy AI

7.2.1 Context and need

Trustworthy AI aims at providing a framework for the development of ML technologies, which guarantees their suitability with respect to Europe's democratic and ethical values. Trustworthy AI is typically divided in four broad dimensions, namely AI robustness, AI explainability, AI fairness, and AI privacy.

Privacy-promoting AI, in the spirit of the EU General Data Protection Regulation can support and empower the streamlining of the EU policy for the protection of citizens' private information. Robust AI techniques help protect intelligent systems from malicious attacks which could cause financial damage or social upheaval, or even compromise critical security sectors. Establishing approaches that evaluate, systematise, and ultimately increase the explainability of AI systems helps build trust between consumers and service providers, but also contribute to social cohesion and the democratic process, by providing a solid foundation upon which socially critical sectors (such as the news sector) can better guarantee the quality of provided information. Algorithmic fairness aims to ensure that the fruits of AI are shared equitably and discrimination/bias against individuals or groups is prevented.

AI4Media brings together the leading edge of Trustworthy AI research in Europe to deliver tools that can assist the ML community in building more robust, fair and explainable AI models for the media industry while simultaneously respecting the privacy rights of EU citizen's data upon which these tools rely on. Concretely, this consists in i) testing the reliability of fact checking AI models against various sources of attacks and building adequate defence systems to reduce the spread of fake news; ii) augmenting classification tools such as Deepfake detectors with explainable algorithms to assist human curators in understanding why a particular piece of media is not genuine, iii) guaranteeing that a fair and balanced set of opinions are expressed in the content recommended to users; and iv) protecting personal data gathered to produce these tools using various obfuscation techniques such as differential privacy.

In the following, we present the main areas of research interest under the Trustworthy AI theme. The aim is to ensure that AI can be trusted to behave ethically, lawfully and offer accurate results.

7.2.2 Areas of research interest

Analysis of legal & ethical frameworks for trusted AI in media environments

→ Research challenges

There are many ethics guidelines both on EU and international level concerning the trustworthiness of AI systems. However, these guidelines do not adequately provide how and in what ways the principles outlined could be applicable to specific sectors, especially concerning the use of AI in media. Media professionals and AI researchers would benefit from an in-depth analysis of these guidelines and from recommendations on how to apply them in media environments.

→ Research directions

Four research lines will be pursued: a) investigation of the principle of "lawfulness, fairness, and transparency" of GDPR to identify uncertainties and gaps with regard to media applications (e.g. the exact limits, scope, and feasibility of the so-called "right to explanation" of the data subjects); b) monitoring of upcoming revisions to legislation relevant to GDPR; c) analysis of the legal data protection framework for the use of AI in media environments aiming to identify areas in which the law could be complemented by self-regulatory measures (including a possible European Digital Media Code of Conduct); d) investigation of the ethics of using big data, its impact on individuals, and safeguards to ensure people are aware of the ways their data is being utilised.

→ Expected impact

Policymakers and the media industry will benefit from the literature review, legal analysis regarding the existing gaps, challenges and opportunities, and policy recommendations based on this analysis. The media industry could also benefit from the practical ethics guidelines that would raise awareness and provide legal certainty on AI use in media.

AI robustness

→ Research challenges

Adversarial AI, aimed at enhancing the robustness of AI models against malicious threats, has been successfully used to highlight vulnerabilities in existing

models as well as increasing their defences against attacks. However, the ability to guarantee the robustness of an AI model to the public is a moving target since new attacks are consistently discovered for which new defence mechanisms must be thought of. Additionally, the ability to communicate to the public the degree to which such a model is safe is an ongoing challenge which needs to be addressed.

→ Research directions

In addition to new attacks continuously arising, the field of AI itself is changing very fast with new types of models, in need of novel defence strategies. Generative AI models, for example, are of particular importance to the media industry. This industry is specifically affected by these types of models both with respect to i) their ability to generate DeepFake content purposely aimed at fooling factchecker models as well as ii) their ability to generate benign illustrations which can be hijacked by attackers to produce inappropriate content. For these reasons, research should also focus on the degree to which existing defences protect generative models from such attacks.

Besides enhancing the robustness of new models, novel approaches must be investigated to establish robustness evaluation methods as part of scalable continuous ML production pipelines. This involves developing adaptive evaluation methods capable of adjusting defence methodologies applied on demand based on the type of attack anticipated.

→ Expected impact

The emergence of generative AI models is of critical importance for the media sector as it promises to provide AI enhanced tools for content producers while also challenging the ability of the industry to identify whether content was genuine or synthesised maliciously as a deepfake. The ability therefore to properly communicate and provide guarantees to the public on the robustness of AI tools being used in this industry will be extremely important.

AI explainability

→ Research challenges

In concert with the growing introduction of AI mechanisms as key decision-making tools in critical domains of our society (e.g. banking, legal, etc.), explainable AI provides individuals with insights about automated decision-making decisions delivered by these previously inscrutable black-box models. Although the number of newly developed interpretability techniques has seen an exponential increase over the past decade, the ability to provide reliable and usable explanations to users supporting a growing variety of AI models remains a major challenge.

→ Research directions

To address AI explainability challenges, we focus on: 1) contributing to the creation of a unified literature covering the exponentially increasing number of newly developed techniques; 2) addressing the need for explainability in novel ML tasks, such as decision-making systems, caption generation, attention mechanisms etc.; 3) evaluating existing explainable methods in terms of their robustness, reliability and impact; 4) evaluating the usability of existing methods with respect to the users along with the types of controls that can be provided.

→ Expected impact

Interpretability stands as a bridge between AI and user social interactions. The need for model reliability (i.e. the generation of explanations that assign meaning to AI decision-making) is thus critical in improving the user's acceptance of a model's decision. For example, regardless of the accuracy of an AI model detecting DeepFake content, any decision must be accompanied by an explanation in order for this decision to be accepted by fact-checkers and journalists.

AI fairness

→ Research challenges

In parallel to the rapid technological development and assimilation of AI technology in every aspect of our lives, the field of AI Fairness has grown recently to address a tendency of some models to exacerbate existing societal inequalities. Although the ability to enhance the performance of AI models so as to make predictions fairer for arbitrary protected attributes has been demonstrated on multiple occasions, these methodologies have yet to be deployed in numerous sub-fields of AI as well as existing deployed models.

→ Research directions

To address AI fairness and AI bias challenges, we will investigate the incorporation of fairness metrics and methodologies within established and novel AI domains such as recommender systems, machine unlearning and federated learning. Models trained within the context of federated learning for instance, by definition are influenced by updates provided by various nodes in a network. This can cause the resulting aggregated models to perform well for dataset attributes held by most nodes but poorly on arbitrary protected attributes of a local node. Introducing fairness metrics directly as part of the algorithms used to produce these aggregate models is thus paramount to protect minority attributes. Finally, bias is addressed in the context of recommender systems, both in the sense of detecting and mitigating respective AI bias and of developing proposals to avoid the reinforcement of human bias (filter bubbles) via recommenders.

The ability to incorporate fairness capabilities to previously trained models is also of significant importance as it removes the need to re-train the increasingly expanding set of models already in use. For this purpose, we explore how domain adaptation neural networks can be modified to answer the question of fairness and bias mitigation in decision systems.

→ Expected impact

Although AI and its benefits are now well established, the potential negative biases of AI systems such as recommender systems or NLP systems towards specific groups or individuals are less understood. Furthermore, the methodologies addressing these biases have yet to be systematically deployed. The ability to incorporate this dimension of AI in model production will guarantee all members of society are treated equally and provided with the same quality of AI support.

AI privacy

→ Research challenges

AI and Big Data have created a need for Privacy-Preserving AI (PPAI), which aims at avoiding input, membership and attribute inference attacks, or model poisoning attacks, applying Privacy Enhancing Technologies (PET) such as Differential Privacy, Homomorphic Encryption, and Federated Learning to achieve their goals. Apart from individual research challenges related to e. g. computational and memory overhead, a core challenge is to adapt and combine PET in an application-specific manner that considers the specific (and often conflicting) application requirements related to privacy, performance, computational overhead, and utility.

→ Research directions

PET are provided as general-purpose tools (e.g. DiffprivLib, Locally Private Graph Neural Networks) and for selected media-specific privacy challenges, including: (i) privacy protection for image classification; (ii) the use of Differential Privacy for recommenders, which may deal with sensitive information such as political preferences, e.g. if news recommendations are based on user feedback; (iii) a proof-of-concept for Secure Federated Learning (SFL) using media content, in which a common AI model can be trained in a decentralised manner, leaving data "on-prem" with the participants, while avoiding inference and model poisoning attacks using Differential Privacy and Homomorphic Encryption.

→ Expected impact

The research aims at solutions that have the potential to "unlock" and improve important media use cases. For example, privacy-aware content recommendation can help media companies / broadcasters to improve user retention and exploit archive material. The SFL prototype can serve as a "template" for common AI training, allowing media companies / broadcasters to exploit assets in a collaborative yet sovereign manner.

AI systems benchmarking

→ Research challenges

Currently there is a growing need for platforms and tools that allow for a higher degree of fairness when comparing different models and methods developed by AI researchers. The automation of the processes involved in creating and maintaining an AI benchmarking competition on a given topic is a vital step in creating a reproducible environment for testing AI models, while also reducing the workload of benchmarking organisers and participants. Furthermore, most of the currently developed platforms only offer tools that measure the accuracy-based performance of AI models, without giving any attention to other important performance

factors, like computational complexity, processing time, hardware resource requirements or task-specific data analysis that could relate to the subjectivity or to the impact of the human factor on the provided data.

→ Research directions

These challenges can be addressed by building a novel platform that creates an Evaluation-as-a-Service environment where interested researchers can upload their data and create, maintain and run their specific benchmarking tasks. This novel platform will encourage task organisers to address the research challenges related by providing them with a set of tools that allows them to easily measure the complexity-related metrics, along with the more traditional accuracy- based metrics, giving a more complete overview of the performance of the proposed AI systems, as well as providing them with tools that allow data visualisation with regard to the subjectivity of the proposed training and testing datasets.

→ Expected impact

These new approaches for the benchmarking of AI systems will allow for a better understanding of the real-world performance of neural models, helping both task organisers and task participants by providing them with more reliable insights into performance and data. Furthermore, in the production stage, interested parties can choose to implement systems by using two-dimensional.

Unique selling point

AI4Media will provide the AI community with a set of novel open source Trustworthy AI toolsets that can be easily integrated within existing AI pipelines and used to enhance the Robustness, Explainability, Privacy and Fairness of models being built by the community. We leverage the technical expertise of AI researchers along with the insights and experience of key media industry players using these technologies in real-life settings.

13 Krizhevsky et al., 2012. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems (pp. 1097-1105)

Accuracy-complexity analysis, instead of just accuracy-based analysis.

7.3 Content-centered AI

7.3.1 Context and need

Content analysis in its various forms has been one of the most widely recognised fields of application for the new generation of AI technologies. Most breakthroughs in AI have become possible thanks to the rise in quantity and availability of large amount of multimedia data[13]. AI systems are asked most of the time to understand highly complex and multimodal data. The huge growth in data availability requires powerful and scalable algorithms to perform summarisation and understanding in multiple domains.

From applications such as object recognition and video segmentation to audio analysis and media synthesis, novel technologies have led to major breakthroughs, opened new markets, and transformed society. AI4Media brings together the world-class expertise available in Europe and participating in the consortium, to advance such technologies and provide foundations for solid use-cases. Technologies that are valuable for the media and news industry, and also marketable as end-user services, such as multimedia metadata extraction, summarisation, and clustering, automatic audiovisual content generation and enhancement, linguistic analysis, and media- specific core technologies to improve learning performance. Within this context, one important focus within AI4Media will also be learning with scarce data and the use of transfer learning, which is key to apply AI when large annotated datasets are not available.

In the following, we summarise the main areas of research interest under the *Content-centered AI* research theme. The aim is to develop novel AI-driven approaches for content analysis and new content creation and address challenges in textual, visual, and audio media and multimedia processing and production.

7.3.2 Areas of research interest

Media analysis

→ Research challenges

The challenge in media analysis is to move from simple annotation of images and videos with class labels – a long-researched topic – to a deeper understanding of the message that a media item can convey. This involves, for instance, jointly

performing more than one analysis tasks; or understanding the events that unfold in a video segment; and, furthermore, providing explanations for these analysis results.

→ Research directions

We will explore various directions that contribute to the goal of deep understanding of media items; these include, for instance, jointly performing optical flow and instance segmentation to improve the results of the latter; exploiting bottom-up information from video, such as detected objects, for recognizing complex events and for grounding the event recognition decisions; and, combining deep learning with symbolic semantic reasoning for the curation of cultural media assets.

Acknowledging the importance of high-quality training data for learning how to perform various analysis tasks, we will also work on the development of novel AI tools for creating, curating and managing media datasets.

→ Expected impact

Media analysis contributes to making non-textual content searchable and findable. This has a profound impact on both the general public – just think of the impact that, similarly, text search engines had and continue to have – and the various media professionals that work with content archives or need to assess if and where a media item had previously appeared (e.g. for media verification purposes).

Media summarisation

→ Research challenges

AI-based media summarisation is a relatively new research area that has quickly gained a lot of momentum. The main challenge is to automatically generate summaries that match the best the user's needs. This requires not only introducing objective criteria, such as the diversity of the generated summary, but also exploiting multimodal data and taking into account the individual user's query or intent.

→ Research directions

The main focus is on the investigation of unsupervised learning methods, i.e., methods that can learn to generate a summary of an input video – be it either in the form of a shorter version of the video or a static summary comprising a number of key-frames – without requiring ground-truth summaries for training. The reason for investigating this direction is that ground-truth summaries are scarce, expensive to produce, and greatly affected by the idiosyncrasy of the human who generates them.

Explainability of video summarisation methods is also investigated, in order to gain a better understanding of how complex AI architectures for summarisation work internally. This will provide useful insights for extending the use of multimodal data – beyond just the visual modality that is traditionally employed – and for performing query-driven summarisation, i.e. exploiting a natural-language description of the user's expectations.

→ Expected impact

Automated media summarisation contributes to easier access to information for all, i.e. helping to consume and understand the gist of a media item in a limited amount of time; and, to the editorial and creative process of distributing content that is suitable for and appealing to its intended audience, e.g. enabling journalists and programme producers to easily generate video highlights.

Media content production

→ Research challenges

Given current advances in AI and deep learning, we expect that, in the future, AI-supported content production will be the norm in any professional media creation pipeline. Open challenges to address include: how to move from simple synthetic content generation to complex video creation; how to support video delivery and automated shooting via trajectory forecasting; how to create high quality content using real-time content restoration.

→ Research directions

To support video archival reuse, we will focus on methodologies able to enhance existing videos; this involves deepening the knowledge of Generative Models such as GANs and deep super resolution architectures and artefact removal. Generative Models are also a fundamental technique to allow the creation of novel visual content via manipulation of existing imagery to the prompt based content generation. Finally, to deliver fully automated

cinematography, core research direction involves the study of a set of forecasting problems among which multiple trajectories prediction.

→ Expected impact

We expect to provide disrupting new technologies for the media production sector. As media professionals are in need of more and more support from AI based tools to compete in an environment where the appetite for new content only increases, delivering new methods for the generation, acquisition and delivery of multimedia content will significantly benefit media organisations, content providers, and users.

Content representation, indexing and retrieval

→ Research challenges

Multimedia content is distributed across modalities (visual, audio, text, etc.), with each modality carrying its own specific piece of information. Interactive media search solutions combining initial text-based retrieval with efficient browsing and exploration functionalities of other media modalities are needed for the development of efficient content retrieval systems for the media sector.

→ Research directions

To exploit all possible modalities together, one of the main challenges is defining how to learn the best representation of input data. Either each modality is associated with a specific representation and is processed separately to be later combined in a fusion schema, or a common global description is processed by a unique general decision model.

Both lines of research should be explored in order to design better information retrieval systems for media data. Considering each modality separately allows dedicated methods which require smaller models. The fusion schema merging all specific modality descriptions into one could allow integrating prior or expert knowledge into the final decision. On the other hand, Transformers provide a new way to process all modalities together in one model, exhibiting impressive performances. However, they require a considerable amount of training data, computation and memory resources while integrating prior knowledge into such models is an open question. New learning paradigms can reduce training time and thus allow fast model update iterations to provide more precise and more relevant retrieval answers.

→ Expected impact

While media companies are the ones producing the content, they currently lack accurate, fast and reliable tools to exploit and thus monetise this content. The media and entertainment industry would directly benefit from next generation search engines that would allow to search: audio-visual archives to support a news story with selected videos; the internet to find content that users like or need; music with specific characteristics to match with film scenes or a textual story maybe matching some of the lyrics; 3D content to find visual assets for a game level, etc.

Language analysis for the news industry

→ Research challenges

Natural language processing methods allow us to characterise news topics and opinions in news platforms and social media. The main challenges in this field are: (1) the ever growing number of new topics and public personalities that emerge in the news and that need to be detected by the algorithms; (2) the fine grained opinions expressed in those documents that need to be accounted for when performing document retrieval.

→ Research directions

The first challenge can be addressed by designing models that can be adapted to a new domain rapidly. We will therefore rely on transfer learning and pretrained language models that have gained a lot of interest during the last couple of years. The second challenge can be tackled via fine grained representation learning that takes into account opinions expressed about all entity mentions in the documents. Named entity recognition can be used to extract topics and entity mentions and on aspect-based sentiment analysis techniques to extract the opinions.

→ Expected impact

Fine-grained news representation can be used in recommendation systems to better filter content and to provide a way of bursting opinion bubbles by allowing end users to select different opinions on the same topic. These representations can also be

used by news providers in their information system to assist them with document retrieval and background research.

Music analysis

→ Research challenges

The challenges of music analysis arise from the inherent subjectivity of music annotations, the use-case-dependence of tag/class taxonomies, as well as the lack of well-annotated large-scale datasets. Similarly, it is hard to train and to evaluate music similarity algorithms based on deep embedding representations due to ill-defined and subjective similarity concepts and low annotator agreement.

→ Research directions

To address these challenges two lines of research will be pursued: a) exploiting domain-specific similarity (rhythmic/melodic/harmonic/timbral) based on conditional similarity networks (CSN), and b) using disentanglement learning for music similarity and training deep audio embeddings, which capture different music concepts such as genre, mood, and instrumentation, and can also be used for music structure analysis, i.e. to identify segments in music recordings, which are homogeneous with respect to one or multiple of the disentangled music properties.

→ Expected impact

Music similarity and structure analysis based on disentangled music concepts allows for an easier integration into already existing expert-annotated music catalogues. Furthermore, concept disentanglement and uncertainty measurement better reflect the subjectivity of users' preferences and will help to increase the acceptance of automatic annotations.

Learning with scarce data

→ Research challenges

Deep learning-based algorithms for multimedia content analysis need a large amount of annotated data for effective training. But in many real-world applications in the media domain, it is not possible or not viable to gather and annotate such a large training data. 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 (e.g., emerging events, locally relevant concepts). Therefore, there is a clear need for novel methods to train models with scarce data.

→ Research directions

There are a variety of approaches to deal with data scarcity, therefore the research directions pursued in AI4Media for this task reflect this diversity. Specifically, we investigate unsupervised and semi-supervised learning (e.g. semi-supervised learning for fine-grained visual categorisation), biologically-inspired learning (e.g. content-based image retrieval with Hebbian Learning), sample-efficient methods (e.g. few-shot object detection in images), domain adaptation (e.g. for traffic density estimation and counting), clustering (e.g. deep clustering with diversity-enforcing constraints) as well as dictionary learning and curriculum learning.

→ Expected impact

The media sector will benefit from novel methods for training with scarce data, as AI techniques can then be applied to problems for which it has not been possible (or economical) to use it before due to the scarcity of data or cost of annotation. E.g., it makes semi-automated tagging of broadcaster archive content doable not only for common object classes (person, vehicle etc.) but also for very specific classes (like a face mask).

Computationally demanding learning

→ Research challenges

Current deep learning approaches generally downsample input data (e.g. images) to manageable sizes (e.g. 200 to 500 pixel images), which while successful may cause critical losses of information. Tasks like, for example, person identification or detection in video, deep fake detection or image and video enhancing are very detail-oriented tasks, highlighting the need for computationally demanding approaches.

→ Research directions

The challenges of AI-based image and video enhancing (super-resolution) can be approached by an exhaustive benchmarking of state-of-the-art AI super-resolution models, both with quantitative metrics and a qualitative human eye test. The

detection of synthetic image and video enhancing can be tackled by training deep learning models with previously synthetically upscaled images and videos. Given the lack of existing training data for this task, we will also work on developing a dataset formed by original and AI-upscaled images at the same resolution.

→ Expected impact

Computationally demanding learning will grant more tools to media outlets to detect, for example, deep fakes and other forms of media tampering, helping fight misinformation. It will also allow more methods of information extraction, for example assisting in person detection and identification in image/video.

Unique selling point

AI4Media will provide novel methodologies for multimedia content analysis, summarisation, and production, spanning all modalities: text, image, video and audio. The developed tools will be easily integrated into existing content-centric AI pipelines in order to allow the academic and industrial communities to develop their media-related use cases.

7.4 Human-centered and Society-centered AI

7.4.1 Context and need

Online platforms are a central source of information for citizens, contributing to the shaping of the public debate and affecting societal practices and norms. The core role of AI technologies in online platforms and media provides both incredible opportunities but also significant challenges. Media AI technologies can be applied to improve political participation, equip citizens against disinformation, and encourage healthy debates and social interaction. Inversely, they can be exploited to misinform citizens, to bias political debates, and ultimately to weaken European democracies. Media organisations and citizens should be equipped to understand the role of AI in the news creation and distribution cycle in order to benefit from the opportunities offered to them while avoiding malevolent usages.

Technologies for multimedia manipulation detection can help stem the tide of disinformation, while automatic analysis algorithms can provide political opinion mining, argument mining, and local news understanding. Multilinguality is particularly important in the European context in order to enable comparative understanding of common points and differences between the way the same topics are presented in various EU countries. User perception measurement algorithms support personalisation but also detect biases and enable "bubble-bursting" by favouring diversity and novelty in recommender systems. The research proposed for this theme builds on the AI4Media advancements on new learning paradigms, trustworthy AI, and multimedia content analysis, and complements them via specific methods oriented toward societal- and human-oriented aspects of AI.

In the following, we summarise the main areas of research interest under the Human-centered and Society-centered AI research theme. The aim is to develop novel methods and tools to better understand the factors underpinning online debate, information production and consumption, and social media experience and put AI to the service of citizens and societies.

7.4.2 Areas of research interest

Content moderation

→ Research challenges

In recent years, the main issue with regard to content moderation has been revolving around determining who should decide which content should be removed, for which reasons, when and how. In this context, the question of 'what should be the model for content moderation: can the problem be addressed through self-regulation (such as codes of practice, codes of conducts), or is there a need for a hard-law EU regulatory instrument?' has been prominent.

→ Research directions

We focus on monitoring and analysing the changing landscape of content moderation, especially in the face of automated content recognition technologies powered with AI tools, exploring how algorithmic content moderation could challenge freedom of expression and other fundamental human rights. This involves a comprehensive analysis of the EU regulatory framework on online content moderation,

including the e-Commerce Directive, the Digital Services Act, the Digital Markets Act, the Audio-visual Media Service Directive, and the Copyright in the Digital Single Market Directive and also an overview of future research directions on content moderation. Based on this analysis, guidelines and policy recommendations will be provided for the development of content moderation algorithms, the design of content moderation policies, and the formulation of regulations that respect fundamental rights without limiting the public debate.

→ Expected impact

Along with our policy recommendations to address media AI challenges, this research is expected to provide clear guidelines for content moderation to AI researchers and media companies. Another expected impact is to create awareness in policymakers regarding fundamental rights challenges and other relevant challenges that strict content moderation policies pose, especially when the use of AI tools for content moderation is implicitly or explicitly imposed in regulations.

Manipulation and synthetic content detection

→ Research challenges

The continuous refinement of Generative Adversarial Networks (GANs) and the recent advent of diffusion-based generative models and Neural Radiance Fields (NeRF) have led to versatile media generation capabilities of unprecedented quality and realism. The continuous appearance of improved generative models and media post-processing pipelines make it extremely challenging to build universal detection methods that can quickly adapt to new developments and can handle real-world challenges and adversarial settings.

→ Research directions

One of the most important challenges pertains to addressing the problem in a continuous learning setting under the assumption that generative models will be continuously improving and the detection models will need to adapt accordingly and timely. Moreover, a big challenge stems from the susceptibility of detection models to adversarial attacks and post-editing operations calling for increasing the robustness of deepfake detection models. The above two directions should also be approached from a multimodal perspective, i.e. the combined analysis of audio and video content.

As a means to tackle the scale of the problem, deepfake detection models will also need to considerably improve in terms of efficiency and compute performance, with the ultimate goal of making it possible to detect synthetic media on end users' devices. To increase users' trust on deepfake detection, much progress is also needed on the front of deepfake detection explainability, fairness and transparency, i.e. tool outputs should be understood by media professionals and should not discriminate against specific demographic groups.

→ Expected impact

The proposed research will equip the media sector with capable and trustworthy synthetic media detection tools and will ensure their fact-checking readiness and capacity. Being able to rapidly verify the authenticity of online media is key in times of crisis or during key events (e.g. wars, natural disasters, elections) when the scale and sophistication of disinformation campaigns increases. Trust in media content is crucial, else we run the risk of turning into a zero- trust society, doubting even solid evidence and facts.

Content recommendation systems

→ Research challenges

Recommenders are a powerful tool for media applications, but their development and use increasingly requires addressing trust aspects, including (i) privacy, especially if sensitive information is processed or can be inferred; (ii) transparency, allowing users to understand the inner working of recommendation, especially regarding the processing of user data; (iii) bias, both in the sense of detecting and mitigating problematic AI biases (e.g. sample bias), but also in the sense of the interaction between recommenders and human biases, especially confirmation bias, which can lead to filter bubbles.

→ Research directions

Differential Privacy will be used for recommender systems to protect user/usage data, aiming at news recommendation as an example where sensitive information regarding political preferences can be inferred. Another important line of work will focus on

detecting and mitigating bias in recommenders, including addressing potential socio-demographic bias in GANs, providing new formalisations for fairness and bias, and proposing respective benchmarks and evaluations. Moreover, proposals on how to deal with filter bubbles that may evolve from recommenders enforcing human bias (confirmation bias) will be developed, by leveraging evaluation metrics for diversity, novelty and serendipity, in addition to standard utility metrics. Finally, improvements regarding explainability of recommenders will be a key element that promotes several of the aforementioned aspects, especially in the sense of transparency and control of how user data is used and processed.

→ Expected impact

Detecting and mitigating biases in recommenders and improving privacy and transparency addresses key requirements from media companies, which means that providing relevant solutions will significantly improve technology adoption in the media domain. In addition, the provision of insights into how recommenders can be designed and evaluated in a way that is less likely to reinforce confirmation bias is a step toward addressing a problem that already has strong implications for political discourse and democratic processes.

Online political debate analysis

→ Research challenges

The main challenge when analysing online political debate is the lack of ground truth, because i) it is impractical to label the amount of data that is being processed, ii) human behaviour is not easily categorizable, iii) misinformation labelling may not be persistent throughout time, and iv) there are no standard metrics to evaluate the healthiness of discussions.

→ Research directions

The healthiness of online discussions on social media can be evaluated by developing a set of discussion health metrics for Twitter debate like measuring discussion ephemerality, estimating the number of bots and bot-generated content, and also monitoring topic-wise public opinion and sentiment analysis. Additionally, the external video content of the tweets will be studied for deep fake presence, and the influence of Twitter discussions on media and vice versa for specific topics will be investigated. To this end, a dataset containing Twitter discussions will be created focusing on various Covid-related topics.

→ Expected impact

Political debate analysis tools can be integrated in existing media platforms and tools, allowing journalists and media outlets to have access to discussion health metrics (e.g. offensiveness, amount of bots, ephemerality). This in turn will allow them to report more trustworthy information and have a better understanding of public opinion and polarisation.

Analysis of hyper-local news

→ Research challenges

Local newspapers are one of the key sources of information for citizens, who are concerned about concrete issues affecting their everyday lives. In Europe, there are hundreds of local news sources over many languages. Many of these local news organisations face challenges on how to adapt to technological trends like AI due to economic and cultural issues.

→ Research directions

AI4Media aims to advance the understanding of the local news information ecosystem across European countries, and develop technology that accurately analyses local news. To identify patterns in news treatment across countries and topics, data is collected from online newspapers across multiple European countries and NLP methods are used to characterise main local actors, themes, and sentiment expressed about specific topics of value for citizens, like for example the attitudes towards Covid-19 vaccination. Furthermore, the problem of how news stories are framed is studied, and machine learning methods are developed to detect specific news framing categories, adapting methods proposed in media studies.

→ Expected impact

Understanding how news sources present and discuss local news plays a role towards promoting information of quality of local value. In this sense, this research will provide a point of comparison to other news sources on topics with significant societal impact like health, where misinformation can result in increased risks and negative outcomes for citizens.

User perception of social media

→ Research challenges
The main challenge in the analysis of social media data is represented by the inherent subjectivity of this particular domain and of the concepts included in it, like interestingness, aesthetic appeal, memorability and emotional impact. This not only creates uncertainty in the training data, but also leads to performances that are lower compared with other domains.

→ Research directions
We will explore several avenues for mitigating this problem. One of the most important methods of increasing system performance, and therefore system reliability, is the use of early and late fusion systems. By using the predictive power of more than one individual system, we can ensure that data processing tackles both the multi-modality of the social media data and that each modality is processed by one or more specialised and targeted systems. While it is important to acknowledge that fusion systems usually involve more processing needs, we will also focus our research on fusion optimisation, attempting to reduce the computation requirements of the fusion systems.

→ Expected impact
The analysis of social media data and its impact on viewers via AI methods can create a useful set of attributes that can be attached to each media sample (is it memorable? is it emotional?). This can allow for a faster navigation of this data and for better results to queries, as well as provide tools that can help content creators better understand the impact of the content they generate.

Real-life effects of private content sharing

→ Research challenges
When posting content online, users are aware of the sharing context. However, this content is then analysed using AI methods, and results can be used in contexts which were initially unforeseen, with a different interpretation. Citizens are entitled to know how their personal data are used, particularly when inferences have an effect on impactful real-life situations.

→ Research directions
The starting hypothesis is that user awareness about data sharing can be raised by providing understandable feedback about the potential effects of automatic inferences. Focus is on the impact of personal photos sharing since they are pervasive in online platforms. A combination of automatic visual object detection and photographic profiles rating and ranking is proposed to implement user feedback. User profiles are rated in impactful real-life situations, i.e. searching for a bank loan, an accommodation or a job, by comparison to a large set of user profiles which were also rated in the same situations. All inferences are performed on the users' mobile devices in order to preserve their privacy.

→ Expected impact
This research raises user awareness regarding the use of their data by online platforms, and by associated third parties. This is important from a societal point of view since informed users will share their personal data in a more responsible manner. Importantly, the practical impact of such research would be amplified if similar solutions would be adopted by online platforms.

Unique selling point

AI4Media will provide citizens and media professionals with a set of AI technologies that can be used to counter the effects of media manipulation, the understanding of online debates, the analysis of perceptions of social media, and the effects of data sharing. These technologies will be integrated in real-life use cases via a tight collaboration between research teams and media industry professionals. The proposed technologies will be open sourced in order to maximize their societal impact.

7.5 Expected outcomes

- → **Open access scientific publications** on topics related to the four research themes
- → **Open access software** for machine learning, multimedia content analysis and trustworthy AI
- → **Open datasets** for media AI research

8. Generative AI and LLMs for the Media

8.1 Context and need

Content creation is the pillar of the media and entertainment industry, with media companies, broadcasters, journalists, artists, influencers, etc. making a living out of producing a large stream of multimedia data. While most of the professional media content is still manually edited and originated, the rise of Generative AI (GenAI), mainly driven by the recent explosion of Large Language Models (LLMs), is already drastically transforming the content creation process for both media companies and independent creators by automating tasks that require creativity.

Trained with massive amounts of data, Generative AI models are able to automatically produce new, original high-quality content in the form of text, image, audio, video, 3D representations, etc. that looks like the data the model was trained on. Large language models like GPT are essentially next-word prediction engines that process natural language inputs, known as prompts, and produce realistic text. Text-to-image models like DALL·E 3 can generate images that accurately match complex and nuanced text prompts. Vision language models (VLMs) like CLIP can learn simultaneously from images and text and are used in tasks like visual question answering, image search or image captioning.

The launch of ChatGPT in 2022 was a turning point that highlighted how GenAI can be mainstreamed to help both media companies and consumers to create new content. Since then, the wide adoption of LLMs and the recent rise of large multimodal models (LMMs) that can process and understand multiple types of data modalities, are creating amazing opportunities for the media industry. The range of potential applications is limitless: deepfakes for film/TV, film preservation, automatic news article production, music composition, realistic game asset development, automatic script generation, personalised advertisements, interactive storytelling, interactive virtual worlds etc.

At the same time, Generative AI presents significant challenges and risks for media companies, creators and consumers but also for the general public. Dataset biases, soaring environmental costs, lack of explainability for model outputs, intellectual property violation concerns, deepfakes and disinformation, AI hallucinations, increasing power imbalance between large media companies / tech providers and smaller companies / independent AI developers and content creators, broader societal implications etc. are only some of the concerns that have emerged recently.

In the following, we summarise the main areas of research interest under the Generative AI and LLMs research theme. The aim is to efficiently address current technical or application limitations of GenAI for the media but also deal with critical ethical and societal concerns. We explore i) legal, ethical and societal aspects, ii) key technology aspects, and iii) media and creative industry applications of GenAI.

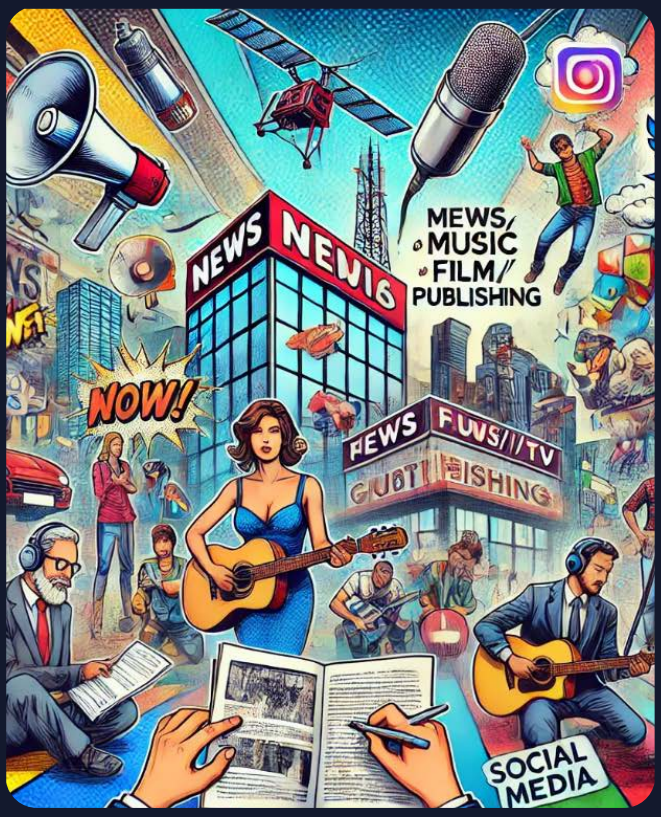


Figure 6: Image generated by Dall-E based on the following prompt: Paint the media industry, including news, music, film/TV, publishing, games, advertising, social media, etc. using a comic novel style.

Verily, the advent of Generative AI doth cast a shadow upon the media realm, fraught with perils most dire. This sorcerous art, capable of conjuring images, texts, and sounds from the void, doth blur the line 'twixt truth and falsehood, making it a Herculean task for the common man to discern the verity of that which he consumes. Lo, the specter of misinformation doth loom large, as deepfakes and other chimerical creations spread like wildfire, sowing discord and mistrust. Furthermore, the quills of many a noble scribe may be laid to rest, supplanted by these unfeeling machinations, thus depriving the world of the rich tapestry of human thought and discourse. In this brave new world, the integrity of our chronicles hangs in the balance, as we grapple with the manifold repercussions of this burgeoning sorcery.

Figure 7: Text generated by ChatGPT using the following prompt: Write one paragraph about the risks of Generative AI for the media industry, in the style of Shakespeare.

8.2 Research on legal, ethical and societal aspects

8.2.1 Generative AI and the AI Act

→ Challenge

There are numerous obligations in the AI Act both for the providers and the deployers of general-purpose AI systems (foundation models) as well as some transparency rules for deepfakes. For media organizations in particular, it is not always clear what the exact scope of these obligations is and how to implement them in practice.

→ Research directions

Chapter V of the AI Act establishes an entirely new framework for general-purpose AI models. Article 50(4) sets a general obligation for deployers of AI systems (e.g. media) that generate or manipulate image, audio or video content constituting a deep fake, to disclose that the content has been artificially generated or manipulated. Exceptions apply such as where the AI-generated text has undergone a process of human review or editorial control and where a natural or legal person holds editorial responsibility for the publication of the content. This provision is particularly important in the context of media. The scope of this exception requires further research and clarification. On the other hand, the lack of explicit legal obligation for transparency does not mean that media do not hold moral obligation that stems from responsible media and journalistic practices and values. The relationship between the two needs to be clarified. Research will be needed to explore the impact of the new transparency rules and the alignment between the Code of Practice on Disinformation, the Digital Services Act (DSA) and the AI Act.

→ Expected impact

Exploring the applicability of these transparency rules for GenAI models and deepfakes in the media sector will bring legal certainty for media professionals and for end-users. Meaningful transparency will support a well-informed media ecosystem, ultimately reinforcing trust.

8.2.2 Dataset diversity and quality

→ Challenge

Challenges related to the creation of high-quality datasets (including high costs, biases and restrictions for data sharing) are limiting the diversity of datasets available for AI model training in the media sector and have led to over-representation of modern Anglo-American content. This creates harmful inaccuracies that are perpetuated in the AI generated content used by media organisations.

→ Research directions

New interdisciplinary research initiatives should focus on addressing linguistic limitations, historical accuracy and cultural biases in datasets that currently hinder the applicability and accuracy of AI models in media organisations across Europe. Particular focus should be paid to non-English-speaking communities and underrepresented cultures, and to ensuring that they are not only being included but rather are leading such research projects. This should be accompanied by research into new digitisation frameworks, AI training methodologies and data governance models that help to overcome dataset limitations while respecting and protecting the values and rights of the represented communities.

→ Expected impact

Research findings will enable the development of tailored algorithms and pre-processing techniques to mitigate linguistic, historical and cultural biases. This will increase the number of media organisations around Europe that are able to benefit from AI in their local contexts, thus improving their competitive advantage and enriching the media ecosystem with more culturally-sensitive and diverse content.

8.2.3 Challenging power dynamics and infrastructure capture

→ Challenge

As the power of big tech consolidates, there several bottlenecks and barriers to entry that make it difficult to enter the AI stack. At the data level, data centers have proprietary access to datasets and therefore hold a competitive edge. The computing power needed to train and host large-scale models for Generative AI is equally held by the few largest private cloud computing providers. It is feared that the gatekeeping role of a few players with a dominant position on the market could be further exploited. The power imbalance between AI developers and providers on one hand and users (e.g. media) on the other is ultimately linked to the so-called "infrastructure capture"[14]. The power concentration could impact media pluralism, media independence, press freedom, protection of journalistic sources, etc.

→ Research directions

First, research is needed to strengthen the digital commons and open public stack for media as an alternative to existing market-driven AI monopolies[15]. This includes investigation into digital infrastructures that are designed to support openness, interoperability, public values and participatory governance.[16] For instance, research is needed into how to build LLMs and other Large models in a way that they protect media values and do not produce ethical trade-offs for media organisations. Research initiatives should define the characteristics and parameters of public value-driven AI infrastructures for media and consider what conditions should be put in place to incentivise both public and private actors to participate. Second,

14 the creation of 'circumstances in which a scrutinizing body [the news media] is incapable of operating sustainably without the physical or digital resources and services provided by the businesses it oversees and is therefore dependent on them [the platform companies]' (Nechushtai Citation2018). Nechushtai, Efrat. 2018. "Could Digital Platforms Capture the Media through Infrastructure?" Journalism 19 (8): 1043–1058. [Crossref], [Web of Science®], [Google Scholar]

15 https://publicstack.net/

16 Signs of progress: Digital Public Infrastructure is gaining traction

research into the solutions that aim to support organisational capacity development in small media companies and weakened geographies is needed.

→ Expected impact

Research on the power dynamics and infrastructure capture in AI and GenAI is crucial for promoting AI democratization and mitigating the AI divide. By understanding the power dynamic and stakes of AI development and deployment, strategies can be developed to ensure fair competition and equitable access to AI technologies. This, in turn, supports a diverse and independent media sector, essential for maintaining plurality and independence. Such efforts contribute to a healthy democracy by fostering informed public discourse and preventing monopolistic control over data and information. The commons-based AI infrastructures for the media sector could serve as an example to follow for other sectors.

8.2.4 AI-generated content and copyright laws

→ Challenge

Generative AI introduces new challenges for copyright laws, both when it comes to copyright-protected input and output. The main challenge for the media and audiovisual sector with regard to copyright and Generative AI is whether or not news content, music, video, art etc. should be used as training data for large models. Many journalists, novelists, artists, and content creators worry their creative work will be exploited by private tech companies building large models. Different jurisdictions have diverging approaches to text and data mining (TDM): the US' permissive "fair use" doctrine vis-à-vis EU TDM exception, which gives rightsholders the ability to opt-out (and demand compensation) of the use of their works by commercial developers.

→ Research directions

Firstly, it is necessary to research the legal and moral concerns regarding the large-scale exploitation of training data such as the knowledge, authorisation, acknowledgement or compensation of their creators. This triggers additional issues of journalistic integrity and accountability to protect journalistic work, as well as what compensation could be considered "fair" so that it accurately reflects the so-called future "transfer value" of the data. For artists, authors, content creators, the most pressing legal issue arises around personality rights and the right to control the commercial use of a person's appearance, artistic expression or voice.

Secondly, it is important to explore the ideas of explicit legal protection against unauthorized use of works for ML training and the introduction of statutory licenses. In the context of media, an important aspect of future research is the impact that some of the deals between the largest media outlets (Financial Times, Axel Springer) and LLMs of other GenAI model developers, have on smaller publishers, content creators and the media ecosystem more generally.

→ Expected impact

The copyright considerations associated with GenAI are raising numerous ethical and legal questions. It is crucial to bring scrutiny and certainty to these aspects in order to ensure a level playing field between the largest media industry organisations and small and local media organizations. It will also ensure certainty for independent content creators whoever they are for their copyright protected work but also solving interrogations for the status of the GenAI output. Providing in depth analysis of these questions will improve fairness in copyright negotiations and bargaining powers and will stimulate creativity.

8.2.5 Impact on epistemic welfare and the (dis)information ecosystem

→ Challenge

Generative AI has the ability to distort the public's trust that what they see is true. This growing anthropomorphization of AI has a profound impact on people's epistemic welfare - understood as the individuals' right to know and be exposed to trustworthy, independent and diverse information.

→ Research directions

Research on the impact of GenAI on disinformation is ongoing and needs to be further supported. The mid-term and long-term impacts should also be included in research plans. The implementation of the DSA's data access requests and transparency obligations enforcement will be key to help researchers gather more evidence on AI's impact on disinformation. Research should also focus on the development of robust methodologies for DSA systemic risk assessments, particularly focusing on the role of AI in spreading disinformation but also on the impact of the presence of AI generated or manipulated content on platforms.

→ Expected impact

Disinformation brings confusion and deception, it is an extremely challenging topic to study and address. Therefore, evidence and facts need more than ever to be gathered. Timely and continuous research efforts are necessary to identify emerging trends, assess the effectiveness of current interventions such as the DSA, the revised Code of Practice on Disinformation, and other instruments. Ensuring a good complementarity between the relevant legal frameworks is of great importance.

8.2.6 Impact of Generative AI in media production and creativity

→ Challenge

GenAI makes it possible to generate audiovisual content for new productions without having to license materials. This might free up time and resources for creators, leading to the emergence of more diverse media makers and productions. At the same time, it detaches new productions from the media value chain that relies on trustworthy, contextualised and ethically produced content and puts at risk business models that rely on co-produced or licensed content.

→ Research directions

Research should investigate how GenAI tools influence creativity, collaboration dynamics, and the artistic vision of creators, with a focus on quality, originality, and authenticity of AI-generated content. This includes analysing media outputs with significant societal influence, such as news, to understand the broader implications of GenAI technology.

Consumer-focused media studies research will provide valuable insights into how GenAI impacts consumer perceptions, behaviours, and preferences. Studying changes in viewing habits, content discovery methods, and platform preferences driven by the integration of AI technology in media production will, for instance, shine light on changing media consumption patterns. Additionally, understanding the economic implications of GenAI on business models and revenue streams in the media industry is crucial - including the footage sales market, market dynamics and distribution of value along the production chain.

→ Expected impact

For policy makers, research findings will serve as a foundation for developing regulatory frameworks and guidelines on issues such as copyright protection, content authenticity, and the role of non-EU players operating in the European Digital Single Market.

Research will offer industry stakeholders a deeper understanding of how GenAI technology affects various aspects of media production, enabling them to make informed decisions regarding its integration into their workflows.

8.2.7 New displacement patterns and hidden labour

→ Challenge

Generative AI has shown to be dependent on the 'hidden' labour of low-paid annotators as well as on being trained on online content without any payment. Furthermore, Generative AI is also reinvigorating fears of displacement amongst certain professional groups. The recent Hollywood strike was a clear example of such fears in the film and entertainment industry. Similar fears have also been detected in the field of journalism and other forms of content production (e.g., advertisement or copy-writing). While it remains to be seen how big an impact GenAI will have on these professional groups, it is clear that there needs to be a better understanding of the relevant societal implications.

→ Research directions

Research has already begun to address questions of the labour conditions of annotators and the production of new displacement patterns as a result of AI and now GenAI. However, more knowledge is needed to grasp which professions are at risk of being displaced and how to ensure there are regulatory mechanisms in place to support those who are displaced. Furthermore, it will be important to understand to what degree content is being 'crawled' and used for model training and investigate how to support media organisations in protecting their data, in order to ensure that the labour of those who produce the content is fairly compensated and thus mitigate exploitation of workers. Furthermore, more knowledge of how to regulate these challenges will also be necessary.

→ Expected impact

This work would support the media industry in providing more insights into the labour-related risks they face, but also support legislative work aimed at mitigating these risks.

8.2.8 Good practices for the use of GenAI in the media sector

→ Challenge

Many media companies are already adopting GenAI tools into their workflows. The relative novelty of these technologies means that there is a vacuum of widely adopted methodology or organisational strategies and policies on how to engage with them in low-risk, ethically- and legally-sound ways. In media organisations - and particularly those in a public service role - this is producing new governance challenges because it can be difficult to ensure compliance across internal and external producers of content.

→ Research directions

Research is needed to develop good practices around the adoption of GenAI in diverse operational contexts in media organisations. This should include using sandboxes to develop case studies around both low risk and high-risk scenarios. These scenarios should consider to what extent existing legal frameworks (namely, the AI Act and the DSA) provide clarity regarding risks, trustworthiness and accountability, and identify gaps where media organisations need to introduce specific policies. Some research has also already begun to explore how media organisations are developing guidelines to ensure that GenAI is used responsibly but more work is needed to understand where gaps remain.

Insights from such research should be used to propose organisational policies, introduce methods to ensure transparency about the use of GenAI, and develop ethical and risk assessments. The AI Media Observatory could serve as an environment to disseminate this.

→ Expected impact

Access to a library of good practices and guidelines will enable media companies to efficiently integrate GenAI tools, ensuring high-quality outputs and adherence to ethical standards. It would foster innovation, reduce implementation costs (due to learning from insights from others) and enhance competitive advantage by providing access to proven strategies and solutions. Such research could help media organisations actively work towards a responsible use of AI.

8.2.9 Generative AI Literacy for media professionals and media consumers

→ Challenge

The rapid advancement of Generative AI presents challenges for both media consumers and professionals. Consumers need to understand how GenAI is used in media and how to critically evaluate GenAI outputs. Media professionals like journalists, on the other hand, may be hesitant to adopt GenAI due to a lack of understanding and concerns about job displacement or compromising journalistic integrity.

→ Research directions

To bridge the gap between GenAI and media consumers and professionals, a two-pronged research approach is needed. First, publicly available educational programs can equip consumers with the knowledge to navigate the increasingly GenAI-driven media landscape. These programs should not only explain the core functionalities of GenAI but also delve into its diverse applications across media formats (text, image, video, audio). Furthermore, the programs should explore the potential of GenAI throughout the media production process, while addressing the complexities involved and the risk management strategies employed by media organizations. Finally, these programs should equip consumers to critically evaluate GenAI outputs and understand the potential impact on trust in media, fostering a more balanced perspective on the advantages and disadvantages of this technology.

Second, training programs can empower media professionals to leverage GenAI effectively and ethically. These programs should provide a clear understanding of GenAI capabilities and limitations, dispelling any misconceptions that might hinder adoption. The training should also equip media professionals with the skills to integrate GenAI responsibly into their workflows, potentially through interactive tutorials, workshops, or even LLM-powered learning companions. Additionally, research should explore ways to make GenAI more user-friendly for media professionals like journalists. This includes developing intuitive interfaces that simplify interaction and provide clear explanations for the models' outputs. Finally, establishing clear ethical frameworks for GenAI use in journalism and the media in general is crucial. These frameworks should address issues like bias mitigation, transparency in source attribution, and ensuring human oversight of content generation.

→ Expected impact

By implementing these research directions, we can expect a significant positive impact on both media consumers and professionals. Publicly available GenAI literacy programs will cultivate a more informed and critical media audience, better equipped to evaluate the ever-evolving media landscape. These programs can also foster trust between media organizations and their audiences by providing transparency into GenAI use. Empowering media professionals with GenAI knowledge and fostering a collaborative environment can revolutionise the media landscape. For example, journalists can leverage LLMs to improve efficiency in tasks like data analysis, fact-checking, and research. This frees up time for more in-depth reporting and investigative journalism, ultimately leading to the production of higher quality, more informative content for the public.

8.2.10 Environmental impact

→ Challenge

The environmental costs of GenAI are soaring, raising fears about the impact on the climate crisis. According to recent reports, a search driven by GenAI uses four to five times the energy of a conventional web search while soon large AI models are likely to need as much energy as entire nations[17]. The environmental cost of AI is currently hard to measure but it is widely and adversely localised, predominantly in the Global South. This risk exacerbating the difference of who benefits from AI and who 'pays' the cost (e.g., mining minerals for GPUs or storing waste).

→ Research directions

Research should focus on various directions: a) develop standards to efficiently assess the environmental impact of GenAI and create relevant public reporting frameworks and mechanisms for developers and operators of large AI systems; b) optimise architectures and build models with a much lower carbon footprint while also developing more energy-efficient hardware; c) identify the aforementioned inequities and develop mitigation strategies that move responsibility to those who also receive the benefits of AI, creating better conditions and bargaining power for local affected communities; d) create regulatory frameworks that promote transparency and accountability with regard to the environmental impact of GenAI and explore the provision of incentives for more environmental-friendly GenAI systems; e) propose AI applications that will decarbonise and increase the sustainability of other economic sectors.

→ Expected impact

Steering the focus of AI research towards sustainability, by reducing its environmental impact and/or increasing the sustainability of other sectors, will promote the green usage of the technology and justify its development. Considering the anticipated consequences of the climate crisis, this goal is not optional anymore. Standardised reporting and better accountability will help steer mitigation efforts to areas with the highest yield and contribute to climate justice.

17 Generative AI's environmental costs are soaring — and mostly secret (Nature, 2024)

8.3 Research on key technology aspects

8.3.1 Domain adaptation and fine-tuning of Large Models

→ Challenge

The main challenges in domain adaptation and fine-tuning of Large Multimodal Models (LMMs) lie in the significant computational resources required and in the difficulty of ensuring that the models generalise well across diverse tasks and domains. Target users adopting LMMs include AI researchers, developers, and industries deploying AI for specific applications. Their main needs are efficient fine-tuning processes, reduced computational costs, and improved model performance on specialised tasks.

→ Research directions

One key research direction is developing parameter-efficient fine-tuning techniques, which aim to reduce the number of parameters that need adjustment during fine-tuning. This could significantly lower the computational burden and resource requirements. Another direction involves exploring methods for effective domain adaptation, such as domain-specific pre-training, to enhance model performance on new, unseen domains without extensive re-training. Additionally, integrating continual learning approaches could further improve models' adaptability and efficiency.

New functionalities should include automated tools for model compression and pruning, which can streamline the fine-tuning process. Developing frameworks for better interpretability and

explainability of models will also be crucial, enabling users to understand and trust model decisions in various applications.

→ Expected impact

Advancements in the research of LMMs will significantly benefit the media sector, enabling more accurate content analysis, recommendation systems, and automated content creation. Society will benefit from more efficient AI solutions that can adapt to specific needs, leading to innovations in various sectors. For example, in journalism, improved models could aid in fact-checking and combating misinformation, thereby supporting democracy by promoting informed public discourse.

8.3.2 Incorporating knowledge in foundation models

→ Challenge

LLMs and foundation models more generally rely on scale (dataset, model size) to encode inductive biases to produce seemingly coherent text or images. Two technical limits ensue: the undesirable implicit biases can resist explicit prompts to go against undesirable association of concepts, and the energy consumed to train and use the models in inference. Leveraging external knowledge in such models is a key to address these issues.

→ Research directions

Injecting explicit knowledge to help disentangle concepts systematically co-occurring in the dataset samples to control the reliance on them and have more rationalized approaches to create sustainable models is a key endeavour.

Two research directions should be pursued. The first is to identify how knowledge representations can be best designed to augment large-scale image-text datasets, so as to control, correct and optimize the semantic dimensions we intend the model to capture. The second is designing new models capable of leveraging suitable knowledge representations, to cut down on required training resources but also to control during training and then transparently audit complex and subtle patterns of correlation learned between sensitive dimensions.

→ Expected impact

This research will result in the design of frugal and less biased models, with improved transparency and efficiency. Knowledge-driven and data-driven approaches are combined to address environmental footprint of AI as well as detrimental social impacts where discrimination is perpetuated by large AI models against most vulnerable groups. The media sector is central in this aspect as the most recent literature shows how subtle and harmful biases are reproduced and amplified in AI-generated content on media platforms.

8.3.3 LLM hallucinations and LLM output quality

→ Challenge

LLMs showed their usefulness in a broad range of scenarios. However, while the "good" answers are generally impressive, if things go wrong (e.g., if the LLM starts "hallucinating" facts), it is hard to tell when exactly this happens without human interaction. One way to tackle this is to combine "classical", more deterministic but weaker approaches to safeguard the LLM output for a given use case, or even use different LLMs to check for output errors or try to improve the LLM itself to make hallucinations less likely or quantifiable.

→ Research directions

Two ways to approach the problem are either to observe the internal state of the LLMs for signs of uncertainty or to post-process the output of the LLM with other tools and other LLMs. Tangentially,

datasets with hallucinated and "real" answers can be created to treat the hallucination detection as a second level classification problem.

However, to avoid the somewhat moot task of detecting an "unreliable narrator", which might be harder than the text generation itself (or sometimes impossible), we propose to start the research on text analysis tasks, where checking for the correctness of the output is easier than creating it in the first place, e.g. on certain text transformation tasks, and proceed from there.

→ Expected impact

By getting more control or a qualitative evaluation of the usefulness of LLM-generated responses, the acceptance of the use of such tools will increase in nearly all problem domains. For domains where a certain quality of the responses is crucial (e.g., fact checking and text generation in the media domain or safety critical applications), it will enable the responsible use of LLMs in the first place.

8.3.4 LLM unlearning

→ Challenge

Given the large scale of the datasets needed to train large LLMs, it is infeasible to properly curate the training data, and sensitive or undesirable information may be ingested by the model during training. Re-training from scratch is impractical and has led to the creation of the unlearning discipline where models are modified to "unlearn" undesirable information without retraining. However, the application and understanding of unlearning in LLMs remains limited.

→ Research directions

Unlearning in LLMs has several applications from removing private/sensitive information to removing biases or reducing the generated toxic content. Thus, it is crucial to develop new, scalable, and efficient unlearning approaches based on the data that needs to be removed. An effective unlearning approach must ensure that the requested data/knowledge is removed from the model and maintains good performance on standard language modelling tasks that fall outside the unlearning scope.

Given that no standard metrics exist to verify that an unlearning mechanism was successful, it is crucial to develop new unlearning techniques and metrics that provide guarantees, leading to improved data governance. Moreover, a comprehensive taxonomy should be created considering that different metrics may be relevant based on the unlearning scenario/task.

→ Expected impact

Developing new unlearning methods in LLMs will lead to: compliance with privacy regulations by removing private/sensitive information or copyrighted data; mitigation of biases in LLMs, leading to fairer models; improved model performance by removing an undesirable subset of data; better aligned models by removing intentionally malicious, harmful or toxic data.

8.3.5 Embodied LLMs

→ Challenge

The main challenge for current applications of LLMs in the real-world is their limited ability to handle uncertainties, such as mechanical faults or task parameters changes. Target users include professionals across various media industries and consumers needing complex task assistance, who require LLMs with adaptability to unforeseen circumstances, deep contextual understanding, precise action sequences under uncertainty and safety constraints adherence.

→ Research directions

To address these challenges, research efforts must concentrate on two main areas: improving LLM reasoning abilities and developing new functionalities for enhanced human interaction. Firstly, creating advanced simulation and fine-tuning methodologies is essential to expose LLMs to

diverse scenarios with uncertainties, allowing them to learn and improve over time. Secondly, focus on developing robust interfaces that facilitate efficient communication between LLMs and the real world.

New functionalities need to be developed for more effective human-LLM interaction and collaboration. This involves creating intuitive user interfaces that enable clear communication of complex tasks and preferences, while also training LLMs on a broader range of experiences through multimodal data processing. These efforts will result in systems capable of accurately comprehending instructions, providing explainable actions, and building trust among users for safer deployment in applications.

→ Expected impact

This research can revolutionize various industries by improving their ability to handle real-world scenarios with uncertainties. This will result in enhanced efficiency, productivity, and cost savings across domains in the media sector. For example, a LLM with a strong contextual understanding and ability to handle uncertainties can more efficiently make plans or provide summaries and insights of evolving real-world situations, such as in journalism. The expected impact of these advancements includes faster content production times, and timely and effective problem-solving, potentially useful for writing assistance and quality assurance.

8.3.6 Dataset biases and fairness

→ Challenge

Training datasets have a key role when training large generative models. Current dataset curation approaches emphasize the size and do not give enough attention to the dataset diversity and quality. Post-training analysis shows that models trained on such datasets encode biases and reproduce dataset-induced biases during inference. The challenge is to propose proactive methods for dataset bias mitigation that focus on harmful biases.

→ Research directions

While AI bias is typically studied in relation to protected attributes (e.g. gender, race), a more comprehensive taxonomy of biases should be established and updated by the AI community that moves beyond the legal definition of discrimination. Such a taxonomy should also include a clear assessment of their potential effects and actionable examples of how these biases can be detected automatically. Second, data diversity should be prioritized when building large training datasets, with an automatic characterization of content from objective and subjective perspectives. For instance, beyond ensuring balance in terms of demographic groups, a number of quality-, perspective- and safety-oriented indicators could be leveraged to assess the suitability of content inclusion in training datasets. This is particularly important for potentially polarizing topics, such as politics, and for topics whose description varies across time and space. Third, objective functions that reduce the effects of biases and encourage diversity should be designed and used throughout the training process.

→ Expected impact

Fairness-, diversity- and quality-oriented dataset curation will lead to improved knowledge encoding in large generative models, contributing to generated content that is less biased and harmful, and at the same time more inclusive and diverse. At the same time, curation and filtering introduce new challenges, especially with respect to freedom of expression. Such challenges call for a multi-disciplinary perspective and approach to the problem.

8.3.7 Enhancing the interpretability of Generative AI models

→ Challenge

The challenge for enhancing GenAI interpretability lies in integrating an interactive layer within generative models that integrates both human-understandable and model-learned

concepts. Achieving this requires bridging the gap between human cognitive frameworks and the representations learned by the models. Another challenge is establishing a direct relationship between these concepts and the generative task itself, ensuring that concept manipulation leads to predictable outcomes in the generated content.

→ Research directions

To address these challenges, research should focus on enhancing the architecture of the bottleneck layer to improve interpretability and usability. This includes investigating methods for enforcing orthogonality constraints among concept neurons to reduce redundancy and ensure distinct, non-overlapping representations. Additionally, leveraging LLMs to acquire and integrate human concepts from prompts is crucial. This involves developing advanced loss functions and optimization strategies that balance model flexibility with the incorporation of human concepts, ensuring that both types of concepts coexist and complement each other in the generative process.

→ Expected impact

By increasing trust between humans and generative models, users in the media industry will gain greater control over content creation processes. For instance, artists and content creators can manipulate high-level concepts such as style or object properties, leading to more intuitive and creative workflows.

8.3.8 Explainable vision-language models

→ Challenge

Vision-language models are an invaluable building block for addressing various downstream video understanding and generation tasks. However, they are "black boxes", and this slows down their adoption. Introducing explainable vision-language models (and in general, explainable cross-modal foundation models) is a key challenge.

→ Research directions

VLMs, such as CLIP and GPT-4, can be used in downstream tasks both as zero-shot methods and as components of a trainable task-specific approach. In the latter case, they may be fine-tuned; or, serve as representation generators, i.e., produce embeddings or natural-language responses that are further used for training a task-specific deep network. Explaining the decisions of such VLM-based methods for downstream video understanding tasks therefore calls for research in two major directions: i) Large explainable VLMs: for this, the opaque nature of the VLMs' inner mechanisms, in combination with the large scale of both parameters and training data, raise serious challenges; ii) Explainability of methods combining VLMs and additional downstream-task-oriented learning structures, where the interplay between the two needs to be understood.

→ Expected impact

Increasing the understanding of VLMs and VLM-based methods for downstream video understanding tasks will help recognize their faults, increase their societal acceptance and facilitate their widespread adoption by the media and other industries.

8.3.9 Resource-efficient vision-language models

→ Challenge

Vision-language models (which constitute the majority of multimodal foundation models) are general-purpose AI models which are able to understand and generate content of several modalities (images, videos, 3D etc.) based on both visual and textual inputs. Unfortunately, the multitude of tasks they are able to solve (like image captioning, visual question answering or content synthesis) necessitate huge model capacities, with both the visual and text component of the models having several billions of parameters. This makes training or finetuning of VLMs very

resource-intensive (usually done on multi-GPU clusters) and results in slow inference times, and ultimately increased energy consumption, which often translates into increased CO_2 emissions.

→ Research directions

To enhance the resource efficiency of vision-language models, several strategies can be implemented. Firstly, model compression techniques such as pruning, quantization, and knowledge distillation can be applied to reduce the model size and computational requirements without significantly sacrificing performance. Additionally, leveraging transfer learning approaches enables the reuse of pre-trained model weights, reducing the need for extensive training (from scratch) on large datasets. Furthermore, optimizing the architecture and algorithms of vision-language models to minimize redundant computations and memory usage (e.g. with mechanisms like Flash-Attention) can significantly improve resource efficiency. Employing hardware accelerators specifically designed for deep learning tasks, such as GPUs or TPUs, can also enhance performance while reducing energy consumption. Recently it has also been shown that surrogates of Ensemble models (like mixture-of-experts or mixture-of-depths) models can reduce the inference time without impacting the model capacity.

→ Expected impact

Making vision-language model training and inference more frugal, i.e. considerably reducing the computational and energy requirements for their creation and operation, is expected to lead to considerable savings in terms of energy, carbon footprint and cost. Additionally, making it possible to deploy powerful vision-language models on commodity hardware is expected to democratise the use of AI and reduce dependency on large corporations.

8.3.10 Physics in vision-language models

→ Challenge

VLMs process and generate visual data from text descriptions. While VLMs have made significant progress in recent years, they still face issues with understanding and representing physical concepts, leading to unrealistic and implausible outputs. Key issues include a lack of physical understanding (e.g. gravity, friction, momentum), inconsistent physical behaviour, ignoring physical constraints (e.g. object permanence, material properties), and limited causality understanding.

→ Research directions

Addressing these challenges requires multi-domain collaboration among experts in physics, computer vision, and machine learning to develop accurate physics models. Research should focus on implementing physics-inspired loss functions that penalise unrealistic behaviour, enabling models to learn physically plausible representations. Additionally, creating new datasets that capture various physical phenomena and training VLMs on multiple tasks, such as object detection and physics-based simulation, will enhance their understanding of physical concepts. For the creation of these datasets, open-source physics engines and VLMs can be used. Another key aspect is developing evaluation metrics to assess the physical realism of outputs, paired with Reinforcement Learning from Human Feedback (RLHF) to iteratively refine the model's understanding of these concepts through realistic simulations and continuous human feedback.

→ Expected impact

This research, including the use of open-source tools and human-in-the-loop reinforcement learning, will result in the creation of VLMs that are more realistic and useful across a wide range of applications. For the gaming and entertainment industry this means the development of more

immersive and engaging experiences, driven by realistic physics. For example, newsrooms could leverage the power of VLMs to create accurate and coherent visualisation of events, such as crime scenes, from textual information provided by eyewitnesses and other sources.

8.3.11 Integration of new/multiple modalities in next-generation Large Models

→ Challenge

The value of Large (a.k.a. Foundation) Models was first demonstrated in Language-oriented tasks: Large Language Models excel in tasks such as question answering and verbal human-machine interaction. Vision-Language Models have followed suit, adding the spatial (2D) image modality to the text one and excelling in tasks such as visual question answering. However, extension of this paradigm to additional data dimensions or modalities, e.g. time (w.r.t. video, time-series data) or specialised data types such as health, financial, IoT data, to name a few, is still in its infancy.

→ Research directions

Of the data dimensions that have not yet fully reaped the benefits of Foundation Models, perhaps the most fundamental one is time. Time is well-understood by humans, and inherently important to many tasks and applications: from video understanding, to robot perception and navigation, to monitoring IoT-generated industrial data or how one's health signals evolve, time is the dimension that makes the rest of the data meaningful. Equipping Foundation Models with temporal understanding and reasoning capabilities is a major future research direction. Perpendicular to it and of similar importance is the direction of combining time (and language, 2D visual information) with domain-specific data such as health, financial, IoT ones. At present, general-purpose models trained on language (or also vision) data can make little sense of such specialised inputs.

→ Expected impact

Leveraging the prowess of Large Models in application domains where the temporal dimension and /or domain-specific data such as media, health, financial, or IoT ones are key, can have a transformational effect. For example, to help journalists cover the topic of climate change, understanding how public opinion evolves over time is important. On the same topic, thinking beyond the media sector's needs, the relevant Earth Observation data and measurements are both time-dependent and go beyond the typical data types crunched by present-time Foundation Models; developing a Foundation Model for them can change our level of understanding of climate change and the factors that affect it.

8.3.12 Collection-wise Visual Question Answering

→ Challenge

Current Large Multimodal Models (LMM) and Retrieval-Augmented Generation (RAG) methods are trained on shared data and designed for shared use. They have a hard time answering complex questions that are specific to a personal, potentially large, multimodal collection. For example, given a large collection of personal videos, to answer questions like "What cars have I ever driven?", "What colour was my dress at my graduation?".

→ Research directions

RAG methods answer a question by including in the generation prompt information retrieved from collections not included in the LLM training data. However, the amount of information that is processed is limited as well as the inference that is made on data. It is necessary to find new computationally feasible models that largely expand the generation context, and train new models on the complex visual Question Answering (QA) reasoning tasks we aim for.

Fine-tuning an LMM to a specific collection requires substantial computational resources. The ability to

efficiently fine-tune LMMs to many different and evolving collections is a key aspect in enabling collection-wise visual QA. Dealing with private, personal collections will also require advancements in privacy-preserving technologies, as the need for safeguarding personal data during retrieval and analysis becomes more critical.

→ Expected impact

The development of effective collection-wise visual QA methods will change the way people interact with their personal multimedia data, enhancing the accessibility and usability of personal digital archives. This capability will not only improve personal memory augmentation and digital asset management but also foster new applications in personalised content creation, automated summarization, and personalised assistant technologies.

8.3.13 Generalizable evaluators for high-quality content generation

→ Challenge

Traditional reinforcement learning (RL) methods rely heavily on human expertise and manual specification of reward functions, limiting their adaptability to complex tasks and the expression of nuanced values. This can result in solutions that are optimised for a specific metric while neglecting broader ethical considerations. For instance, RL-based applications for content creation or recommendation might prioritize user engagement, regardless of the quality or diversity of the content produced. Similarly, RL algorithms guiding the generation of news articles titles, could favour excessive sensationalism or clickbait style over factual accuracy and journalistic integrity.

→ Research directions

One potential research direction is to explore the use of LLMs as a tool for generating and refining target functions in RL. This could involve training LLMs on large datasets of natural language descriptions of tasks or problems, which would enable them to learn about the underlying values that are relevant to each task. The LLMs could then generate target functions that reflect these values, allowing for more effective optimization of the system's behaviour.

Another potential research direction is to investigate the use of multimodal data sources as inputs for LLMs, allowing them to learn from different modalities such as images, audio, or text. This could involve training LLMs on datasets that combine multiple modalities, which would enable them to learn about the relationships between different types of data and evaluate different modalities directly, without requiring manual intervention.

→ Expected impact

By enabling more accurate and human-acceptable definitions of target functions, AI systems can generate content that is more relevant, engaging, and aligned with human values. This also has the potential to increase transparency and accountability through more helpful responses, and broaden applications across various domains such as robotics, healthcare, and education.

8.3.14 Face dataset anonymization

→ Challenge

Machine learning models are trained on datasets that may contain information that can be used to identify individuals. In certain cases, such data cannot be shared to the academic and research community. An important research direction investigates ways to anonymize those datasets by using appropriate generative methods. De-identifying faces can be achieved using many “catastrophic” techniques, such as blurring, masking, or pixelating. Whilst these techniques are usually very effective in protecting the identity of those depicted in facial datasets, they render the data virtually useless in downstream machine learning tasks (such as facial

attribute classification or facial expression recognition). The main challenge in dataset anonymization is the preservation of crucial information in conjunction with privacy protection.

→ Research directions

The main research question that needs to be addressed is how to generate synthetic faces based on real faces depicted in existing datasets, so as i) the privacy of those depicted is protected (i.e., the generated synthetic faces should not lead to individuals depicted in the real dataset), while at the same time ii) crucial characteristics (e.g., facial attributes, facial expression, head pose) of those depicted are preserved.

The aforementioned research goal should be served by devising methods that condition the face generation process on specific information of interest. For instance, one such approach can be based on the facial expression and/or head pose of the real faces, using their facial landmarks, an appropriate generative method (e.g., Diffusion Models), and an appropriate conditioning mechanism (e.g., ControlNet). This will endow the anonymization methods with strong controllability, where synthetic faces will preserve crucial information (e.g., emotion, head pose), whilst at the same time they will convey the identity of a fake/non-existent face, protecting this way the identity of the real face.

→ Expected impact

Dataset anonymization is expected to be of utmost importance both in the AI research in general and in the media sector in particular, since it allows for utilising existing datasets of human faces that at the time of the dataset creation did not give explicit consent in using their identities. Moreover, it allows for collecting new datasets, where humans will not have to reveal their identity through images of their faces, but can still contribute to crucial data collections through non-identifiable characteristics.

8.3.15 Enterprise context-aware Generative AI

→ Challenge

Although framed in their national and European legal framework, existing public service media businesses are bound to a collection of additional principles (regulatory, ethical, mission-related, legal) that are normally expressed through a series of written contractual provisions. The challenge is that of making sure that the adoption of a foundational model or of an application-specific one does not incur in the violation of one or more principles expressed in internal regulations, or that there are enough robust means to assess this risk in advance.

→ Research directions

The main critical aspect is the complexity of internal regulations and how these relate to different departments and processes of a media company, as well as the deep referential structure of some regulations, based on existing (and ever evolving) national laws and on de facto procedures and processes. One research direction includes the development of methodologies to generate test suites able to assess infringement risks. This encompasses not only procedural and/or algorithmic inventions, but also methodologies to access, summarise and include contextual laws, procedures and knowledge bases and use them to formulate test queries that are specifically aimed at the evaluation of this kind of risk. Another research direction could be that of developing approaches to lower/minimise risks in given situations, e.g. through online prompt optimisation/enrichment up to full-fledged finetuning approaches.

→ Expected impact

This research would streamline the adoption of LLMs in the media industry practice, by giving concrete means to objectively assess risks and implement mitigation plans. Depending on the actual dimension of the impacted business, which may be involving

affine sectors to media like e.g., culture/educational, institutional, the same results could represent a driving factor towards the AI industry.

8.4 Media and creative industry applications

8.4.1 Personalised LLMs for journalists

→ Challenge

The main challenges in developing personalised LLMs for individual journalists include sourcing and curating tailored training data, fine-tuning pre-trained models to match the unique writing styles and preferences of different journalists, and ensuring topic specialisation while maintaining interpretability and transparency. The target users are journalists and media professionals who need efficient, accurate, and adaptable tools to enhance their writing and reporting processes. Their main needs include personalised assistance that respects their unique style, improved efficiency in content creation, and reliable topic-specific insights.

→ Research directions

Research should focus on creating methodologies for collecting and curating high-quality, journalist-specific training data that can effectively capture individual writing styles and preferences. This involves developing automated tools for analysing and extracting style patterns from a journalist's body of work. Additionally, techniques for fine-tuning LLMs to incorporate these styles without sacrificing the model's versatility and generalisation capabilities are essential. Exploring methods for achieving high topic specialisation within these models can ensure that journalists receive contextually relevant and accurate assistance.

Another crucial research direction is enhancing the interpretability and transparency of personalised LLMs. Developing frameworks that allow journalists to understand and trust the AI's suggestions and outputs is vital. This includes creating clear, explainable AI mechanisms and interfaces that provide insights into how the model generates specific recommendations. Allowing end-users to tweak and further fine-tune the model is also important. Finally, interdisciplinary collaboration between AI researchers, linguists, and journalism experts can ensure the models are both technically sound and practically valuable.

→ Expected impact

This research can impact the media sector by significantly improving the efficiency and quality of journalistic work. Personalised LLMs can help journalists produce more accurate, engaging, and timely content, enhancing their productivity and reducing workload. For society, this means access to better-quality news and more in-depth reporting, contributing to a well-informed public. For example, a journalist specialising in environmental issues could use a tailored LLM to quickly generate detailed, stylistically consistent reports, ultimately fostering greater public awareness and understanding of critical topics.

8.4.2 LLMs as an assistant research tool for media analysis

→ Challenge

Integrating LLMs into media analysis workflows presents significant challenges, primarily balancing the innovative potential of LLMs with ethical and reliability concerns. Ensuring the confidentiality of sensitive data and validating the relevance of LLM outputs are key challenges, especially for complex textual analysis. Additionally, the high resource consumption of these models is a concern. In this consideration, the target users, including media researchers and academic scholars, aim to enhance the depth of their analysis and their research scale, which differs from the immediate, often time-sensitive and domain-restricted, needs of journalists. Moreover, the epistemological considerations for researchers involve a deeper engagement with the

theoretical frameworks and methodologies that underpin media studies, focusing on the validity, reliability, and reproducibility of findings, which contrasts with the more pragmatic and outcome-focused approach in journalism.

→ Research directions

Building on the idea of utilising LLMs for media analysis as foundation models for text comprehension, the main research directions should focus on developing methodologies for integrating these models into research workflows effectively and ethically. This includes creating frameworks for the automated analysis of textual data that can identify biases, ensure the accuracy of complex inferences, and maintain the integrity of the source material.

Additionally, there is a need to develop new functionalities that enable real-time suggestions for questions during sociological interviews and content exploration, all while embedding mechanisms to safeguard against borderline and harmful generations, for participants notably. These advancements will not only automate research processes, enabling larger-scale analysis, but also open avenues for more nuanced and diverse automated analyses, which are key for academic rigour and comprehensive media studies.

→ Expected impact

Incorporating these models into media analysis can enhance academic research by examining news narratives and public discourse more deeply. These tools can help identify areas for critical investigation and relocate the democratic decisions in the public debate. By providing rigorous and comprehensive media analysis, these models may help us better understand the dynamics of power reflected in the media. Unlike journalistic applications, which often focus on immediate reporting and audience engagement, the use of LLMs in media research aims to contribute to long-term understanding and theoretical advancements in media studies and knowledge theory.

8.4.3 Indexing the entire World Wide Web with LLMs embeddings

→ Challenge

Companies such as Hugging Face, Mistral, and Meta continuously develop and release increasingly advanced representation models. These "third-party" models seem to be the only way users can leverage the advancing performance of AI, especially as models, datasets, and computational infrastructure grow in size, complexity, and cost. However, these advancements require frequent and computationally demanding reindexing of media content, posing challenges for both private and public retrieval services. These services must reprocess all content with each new model release, potentially disrupting continuous service.

→ Research directions

The challenge in fully exploiting this standard practice of using third-party models in retrieval and search systems involves addressing the problem of learning compatible representations. This issue arises from the need to align representations from different models that have been trained with varied datasets, initialization seeds, loss functions, or architectures—either individually or in combination. In such applications, maintaining alignment is crucial to minimize the frequent reprocessing of gallery content—such as images, audio, and text—for feature extraction each time a new pre-trained model is introduced.

The ultimate challenge lies in using LLMs to represent the entire World Wide Web, a task complicated by its immense scale and constant updates. In this context, learning compatible representations opens the door to potentially and effectively using Retrieval-Augmented Generation (RAG) at the WWW scale. RAG can dynamically enhance search results by combining traditional retrieval methods with the generative capabilities of large language models.

→ Expected impact

The research into aligning representations making them compatible across various models has significant implications for the media sector and society at large. By standardizing how models retrieve data, this technology may ensure for example that news agencies and media platforms can quickly and accurately access and deliver content, which is crucial in maintaining an informed electorate. During elections, such systems could enhance the accuracy and speed of reporting results and analyses across different platforms, reducing misinformation and fostering a more informed public.

8.4.4 Media agents

→ Challenge

Media business is too complex and articulated to be addressed by a bunch of locally-optimised independent Generative AI models. Although data, knowledge and practices about how to run the business is stratified, diversified and distributed in many repositories, nevertheless it is intensely interconnected. The challenge is to explore to what extent AI agent technology can help increase the efficiency of such diversified media processes.

→ Research directions

One research direction could be studying novel interaction mechanisms and communication paradigms for distributed agents in the media domain. In addition to traditional Q&A approaches, i.e. paradigms in which agents collect and elaborate data to provide an answer to a specific user question, advanced paradigms based on proactive behaviour (e.g., event-triggered repository exploration), based on inter-agent communication and actionable outcomes should be pursued. Based on a complete survey and characterisation of modern digital media business, this research would start from addressing innovation in the following main areas: i) innovative agent problem solving algorithms and paradigms, beyond state-of-the-art (e.g., multi-agent ReACT); ii) innovative inter-agent communication architectures; iii) multi-modal agent paradigms.

→ Expected impact

Results of this research would support the media business with the key innovation of a distributed layer of interacting agents able to provide added value to the mainstream production and publication processes, in the form of e.g., automated information gathering and interlinking, original editorial recommendations, contextual or user data feeds and other kinds of exploitable insights.

8.4.5 Audio-visual synthetic media detection

→ Challenge

Due to the rapid speed of Generative AI model advances, the diversity of model types and providers as well as new multimodal training approaches, there is a continued need for updated audio-visual synthetic media detection technologies and related AI services/components for professionals in media and journalism. Due to rapid developments, it is paramount that detection tools are fully transparent in terms of capability and limitations.

→ Research directions

On one hand, continued research is needed to support the development of synthetic media detection tools that target and reflect the new diversity of Generative AI model types and providers - covering proprietary and open-source models, diverse media types (text, images, video, audio/speech, music, 3D-objects, etc.), specific content output (e.g. human-like avatars, portrait photos, human speech, etc.) and ideally supporting model attribution, i.e. giving information on which tool/model from which provider was used to generate the content item that is subject to analysis.

On the other hand, and in context of fast-moving Generative AI technology markets, research should focus on how essential transparency information can be provided with every detection tool that allows

media managers and end-users (journalists, fact-checkers, or researchers) to judge how up-to-date the detection tool is, what are its limitations, to what extent they can rely on its predicted results, and in which way they can interpret/use the result.

→ Expected impact

Despite technological challenges, synthetic media detection tools remain an essential part of a mix of approaches to counteract disinformation and to sustain a healthy information ecosystem - alongside provenance markers for published content as well as media and Generative AI literacy on the consumers' side.

8.4.6 Content provenance markers

→ Challenge

There are a number of digital content provenance marker solutions/approaches under development, e.g. the C2PA standard from Coalition for Content Provenance and Authenticity, SynthID from Google Deepmind, or ISCC from Liccium. The media industry faces challenges related to implementing such solutions across various media types and services and within established production systems, workflows, and processes.

→ Research directions

One area of research should be the general aspect of standardisation as well as related technical integration frameworks for different workflows, production and archival systems and content formats, potentially multiple standards, and the processing of media content in real-time. The latter can also include verification solutions that can detect tampering and manipulation of provenance markers.

Another research focus should be on how to best communicate the existence, effectiveness, and use of provenance markers by the media industry towards users/audiences once they are integrated.

→ Expected impact

The successful implementation of verifiable content provenance solutions in practical media workflows - including communication to users and audiences - can enhance both trust and credibility in the media and journalism sector. It may also help to counteract disinformation by proving content authenticity, e.g. during elections.

8.4.7 Ensuring transparency and the right to be informed through labelling the use of Generative AI in media production

→ Challenge

The rising use of Generative AI in media production makes it challenging to decide when and how to inform end users about it. As AI becomes more integral to tasks, distinguishing AI-generated from human-generated content will be difficult, potentially causing confusion and mistrust. To address this, there is a need for standardised guidelines for transparent communication, balancing openness with practical media production needs. End users have the right to be informed about AI being used but these standardised guidelines cannot be in the way of the user's media experience.

→ Research directions

Research should focus on the development of transparency standards, laying out clear guidelines for when and how to disclose AI use in media, studying the impact on audience trust, and finding the best ways to inform the public without overwhelming them.

With advancements in conversational AI, people will increasingly interact with machines to consume media. This makes it crucial to study the most effective and user-friendly ways to inform users about AI involvement, ensuring they understand what content is AI-generated.

→ Expected impact

Research into transparency standards will positively impact the media sector by fostering trust and credibility, as audiences will feel more informed and confident about the content they consume. For society and democracy, clear disclosure practices can combat misinformation and ensure that public discourse is based on accurate and trustworthy information. For example, if news outlets consistently disclose AI-generated articles, readers can better assess the content's context and reliability.

8.4.8 Real-time music generation

→ Challenge

An important feature of music manipulation is its use in live conditions, as most performers may use this technology besides the controlled conditions of a music studio. Live use implies real time processing, which for large models may be challenging. Also, the model performance capabilities in these conditions should be similar to those of batch processing.

→ Research directions

Among the potential directions for research into the use of large models in music is the utilization of model-based solutions for sound generation. These models may use a physical description of the sound processing, which is simply described by the few parameters that characterize the model settings.

Additionally, identifying novel features to label and classify music, and the combination of these features in a variety of formats (audio and text, as an example) is currently a major effort in order to improve the performance of current models for music generation.

→ Expected impact

Improved capabilities of audio processing will immediately translate into a collection of tools used by artists, in the same direction as any prior technical development that arrived to the music industry. New sounds will be generated with novel (or integrated) tools in the artists' palette, reaching a large audience judging these developments.

8.4.9 Emotionally expressive avatars

→ Challenge

The quality of an avatar's emotional expressiveness plays an important role in user engagement, narrative impact, and user acceptance and experience in media & entertainment applications. Developing naturally expressive avatars is very challenging due to the complexity of human-like facial expressions. Facial expressions exhibit variations and nuances across individuals. Moreover, achieving a realistic look requires not just a single frame but a sequence of frames with seamless transitions between emotions and their intensities. Additionally, synchronization of facial elements, such as eye blinking and lip movement, must be aligned and potentially coupled with voice and body movements. Finally, shared virtual worlds necessitate understanding the user's emotions and translating them into the avatar based on inputs like control point tracking or speech in real time.

→ Research directions

In the past couple of years, there have been significant advancements in image generation, primarily due to the development of diffusion models. These models can alter almost any visual input to resemble another or create entirely new imaginary scenes and worlds. State-of-the-art models are now capable of creating realistic images with strong and diverse facial expressions. Significant progress has also been made in video generation using diffusion models.

It remains to be explored how these recent technological breakthroughs can be leveraged to create videos of emotionally expressive avatars. In particular, focus should be placed on achieving stable and controllable output throughout an entire clip. Another critical aspect to consider is the optimal balance between output quality, performance requirements, and time-to-create, especially in the context of real-time VR applications.

→ Expected impact

This research can lead to better and more efficient ways to create emotionally expressive, input-responsive avatars in real time. This will enhance user immersion by creating more realistic interactions and emotional connections, enriching the overall user experience. The results could be applied not only in gaming environments but also in broader applications such as education, therapy, and communication.

8.4.10 Multimodal text instructable agents in interactive worlds

→ Challenge

Methods for creating LLM-driven game-playing agents that not only act, but also communicate in a contextualized fashion within games is still an early technical research topic. While the potential is clear, and has been demonstrated in individual games, much work remains to produce more general methods for different game types. Once this approach becomes more widely used, research frameworks are needed to understand the sociopsychological impact of these technologies and how to apply them ethically while supporting and protecting players.

→ Research directions

Research with regard to the technical implementation of such LLM-driven game-playing agents should focus on: generalization of multimodal-LLMs for application across games; optimization and control of LLM-agent behaviours in games using methods such as prompt engineering, fine-tuning and RAG-databases; decreasing LLM latency for maximizing relevance in real-time simulation/game settings, minimising LLM computing costs for production relevant cost levels, and hosting multimodal LLMs on consumer-grade edge devices for cost-effective and user-relevant delivery.

At the same time, the impact on the end user under various scenarios, from the cognitive, affective, and ethical perspectives, should be studied. This includes research on the development of frameworks for understanding and measuring the experience of AI agents in play settings as well as the development of appropriate frameworks for understanding the ethics of blending human and AI-agent interaction in spaces.

→ Expected impact

Fully interactive LLM-driven agents for online multiplayer experiences offer to open up a new category of interactive online play experiences and may allow the media sector to innovate on how audiences interact with existing and new universes. Having research frameworks in place for understanding such experiences before they become commonplace can support a societally beneficial adoption of such experiences and may help guide policy creation on the topic.

9. Use cases in media, society and politics

9.1 Introduction

Seven media-related use cases are explored, informed by emerging market opportunities and urgent industry challenges (Figure 6). They cover a variety of topics such as disinformation, news research and production, media moderation, organisation of audiovisual archives, game design, human-machine artistic co-creation, social science research etc. The use cases aim to address significant challenges currently faced by different media industry sectors and to highlight how AI applies throughout the media industry value chain, from research and content creation to production, distribution, consumption/interaction, performance and quality measurement.

Figure 8: The seven AI4Media use cases

The use cases are realised through close collaboration between AI researchers and media professionals (European media organisations or content related companies). Driven by use case requirements, tools and systems used by media organisations are upgraded with novel AI functionalities resulting from the research activities along the main research themes described in Section 7. The use case systems and platforms that are used as the basis for AI4Media's demonstrators cover a wide range of system types (from internal Content Management Systems and tools to fully-deployed Software-as-a-Service solutions) that address a variety of media processes in different industry sectors. As such, the resulting demonstrators showcase the potential uses of AI in the media industry, including aspects of human-centric, ethical and trustworthy AI. They highlight how AI applies throughout the media and content value chain (Figure 9) and how different types of media players aim to address user and business needs with novel AI solutions.

The use cases are developed in the context of Work Package 8.

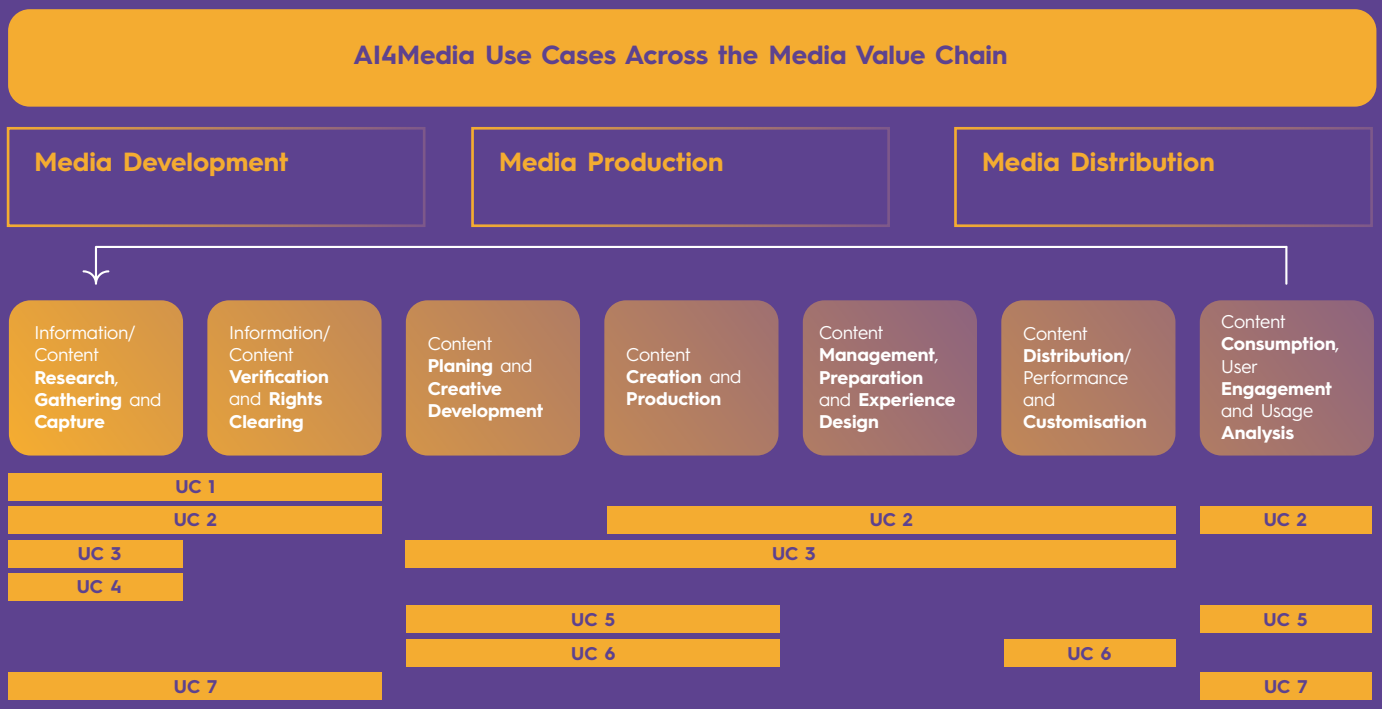


Figure 9: The seven AI4Media use cases and their relevance to the media and content value chain

9.2 AI4Media Use Cases

9.2.1 UC1: AI for Social Media and Against Disinformation

→ Challenges

For the news media industry, it remains a challenge to keep up with increasing volumes of manipulated social media content, synthetic media, and disinformation. Another challenge is related to the dynamic advancement in technologies and techniques used to produce and spread such disinformation related content.

→ User needs

The key actors in counteracting disinformation in the news media industry are content verification specialists, fact-checkers, and journalists. Although they have tools to support this largely manual and complex content verification workflow, there is a need for more specific and AI-powered support services that can be easily integrated into existing systems.

→ Integration of novel AI functionalities

New types of AI functions that can support journalists with the detection, analysis, and verification of disinformation refer to content/account verification and the analysis of disinformation related narratives. This includes AI multi-modal and multi-lingual tools for the detection and verification of synthetic or manipulated media (e.g., text, photos, videos, or audio), analysis of social accounts or tools for identifying patterns and communities in social media discussions. AI tools based on NLP and network science methods could also support journalists and fact-checkers by automatically detecting false claims or flagging suspicious content, detecting cross-platform disinformation campaigns and coordinated inauthentic behaviour networks, predicting the impact/virality of disinformation, and tracking the spread of disinformation across platforms and languages.

→ Expected impact on media, society and democracy

The provision of AI-powered support functions for human decision-making in the content verification workflow can make this complex task more efficient, easier to conduct for non- experts and more impactful. This has the potential to strengthen the role of news media organisations in counteracting disinformation and detecting synthetic media.

Unique selling point

Going beyond stand-alone prototyping, UC1 integrates and tests a set of AI-powered functions in the user context of existing tools for content verification, and the wider business needs of news media companies, including aspects of Trustworthy AI.

9.2.2 UC2: AI for News - The Smart News Assistan

→ Challenges

The fast changing digital news environment poses a significant challenge for the news industry. Journalists experience a lot of pressure as they are required to create trustworthy stories in many different formats in order to be able to respond to the changing habits of media users. This use case focuses on integrating AI-powered tools in journalistic workflows, aiming to reduce the time spent on tedious and repetitive tasks related to news creation.

→ User needs

Journalists are getting their stories from a lot of different sources (social media, mailing, other media, etc.). In order to create relevant and trustworthy stories there is a need for monitoring assistance and factchecking tools. In addition, the fragmented media landscape requires news creators to be present on an increasing amount of diverse (social) media platforms. Since every platform has its own way of storytelling, support in semi-automatically creating different formats of the same news story is needed.

→ Integration of novel AI functionalities

Several new AI-based modules can be integrated to facilitate news creation: tools to perform image forensics including user interfaces for improving understandability of AI results; graph- based network analysis tools to categorise and detect (disinformation) websites and assess URLs coming up during the research phase of writing a news story; tools that based on the story can suggest images or videos to automatically compile new short videos, allowing the journalist to make corrections and additions and do a final edit; tools that provide a diversity score to help write balanced stories; tools that help journalists monitor and follow stories spreading on social media.

→ Expected impact on media, society and democracy

Building a hybrid human-machine news workflow by integrating AI-driven components in critical creative news processes is a complex and not straightforward challenge. By creatingprototypes and making the technology tangible, newsrooms can understand AI functionalities better. That could help them create relevant, trustworthy stories that reach a broader audience.

Unique selling point

UC2 will interweave smart AI-driven tools in day-to-day journalistic workflows aiming to optimise tedious and repetitive tasks involved in news creation and create opportunities for new story formats to reach wider audiences.

9.2.3 UC3: AI in Vision - High quality Video Production and Content Automation

→ Challenges

This use case aims at demonstrating how emerging AI-based technologies can support Public Service Media (PSM) in the transition from its traditional business to the modern digital era. The main challenges to be faced are related to the exploitation of AI's potential to underpin and speed up content production processes and to enhance existing audio-visual content, coming from various sources such as broadcasters' archives and social media.

→ User needs

The main user's needs are related to the exploitation of the immense amount of produced and archived audio-visual content in innovative ways, the need to discover and explore relationships between content and external knowledge, and the need to define and implement innovative ways to produce content. Although substantially targeted at the "Journalism" and "Infotainment" subdomains, this use case represents a good example of the generic integration of AI tools in a modern TV production environment.

→ Integration of novel AI functionalities

The strategic AI functionalities mostly requested in a modern TV production environment are those able to describe, characterise and index content under an as much complete and context- independent manner as possible. This is key in order to catch up with the dynamism and continuous evolution of the media market in terms of new content interaction models and new business domains through efficient means to access and use assets at hand. The work nowadays accomplished by documentalists is still based on the former need to preserve content for archival

purposes, and hence the corresponding annotation and indexing mechanisms suffer from this limitation. Therefore, AI tools that allow to obtain "latent" context-independent descriptions (and corresponding indexing functionalities) make it possible to exploit different takes on available content and feed new applications and services with less effort.

→ Expected impact on media, society and democracy
When a huge patrimony of content like that being preserved and produced by PSM is made exploitable by unprecedented means and unequalled possible perspectives, citizens have certainly a direct benefit in terms of increased, more appropriate, unbiased, and timely knowledge about what's happening around them. As a consequence, the impact on the underlying democracy mechanisms cannot but generate a positive effect.

Unique selling point

Considering the whole production process rather than individual steps in the value chain, UC3 allows to better understand and evaluate the impact of AI technologies at full scale in the typical workflow of a modern TV production environment.

9.2.4 UC4: AI for Social Sciences and Humanities

→ Challenges
A challenge that scholars in social sciences and humanities (SSH) encounter is that manual media investigation methods do not scale across large collections thus limiting their research scope. While various AI solutions are emerging to address this, in many cases they are not tailored to specific SSH research needs or are not considered as transparent or trustworthy enough to enable rigorous investigation and interpretation of the resulting AI analyses.

→ User needs
SSH researchers require tools with easy-to-use interfaces that allow them to approach media content investigation through concepts from media theory such as framing, discourse analysis and narrative theory. This presents an interesting challenge for AI tools as these concepts are largely dependent on the specific context of investigation and are difficult to translate into technical specifications.

→ Integration of novel AI functionalities
Various existing AI techniques can be tailored to enhance SSH research workflows, including: (i) identification of identical audio/video/image elements across multimedia content to track reuse and framing, (ii) NLP techniques to support comparison of opinions, political viewpoints and emotions on a selected topic, (iii) performing queries by text, images or sound to find identical or related segments in audiovisual collections, (iv) use visual and audio clues to identify narrative elements (e.g. conflict, dialogue, etc.). The configurability of parameters in AI tooling is a key consideration for implementation allowing researchers to experiment with different settings, which enables them to better understand and trust the provided results.

→ Expected impact on media, society and democracy
AI tooling for SSH research will help scholars to perform data analysis on both a macro and micro level so that they are better equipped to investigate biases and issues around framing and representation in media. Being better able to identify these issues on a larger scale will help spread awareness and possibly prevent these issues in the future. Also, providing scholars with a better understanding of the requirements for trustworthy AI, the general trust in these tools and therefore their correct use, will improve.

Unique selling point

UC4 not only provides scholars and journalists with the opportunity to use state of the art AI techniques to research broad societal questions on framing, representation and bias, it also hands them the tools to set their own configurations to fulfill their exact needs.

9.2.5 UC5: AI for Games

→ Challenges

When it comes to the use of AI in the game industry, there are two major challenges: i) leveraging and integrating AI methods to support and automate the creation, testing, and verification of modern video games; and ii) allowing practitioners in the video game industry to leverage AI in their production flow in ways that support existing paradigms and interdisciplinary collaboration.

→ User needs

The cost/labour intensity of video game development is increasing due to competition and quality requirements. Practitioners in video game production have a strong need for automating manual and repetitive tasks, e.g. playing through games each time an iteration or content addition is completed. AI methods must be deployed in ways that are readily available, fit in existing workflows in tools for non-AI-expert users, and apply to highly variable game designs.

→ Integration of novel AI functionalities

The game industry can benefit from AI in the following ways: i) integrating AI-based methods for automatic game playing into readily deployable tools, while maintaining generality across varying game designs and interaction schemes, for instance through modern Quality-diversity search algorithms or imitation learning; ii) using the methods to generate play data that characterise games in terms of functionality and performance, reducing the need for manual labour.

→ Expected impact on media, society and democracy

The successful application of AI in video game production will allow for more sustainable and efficient game development processes in the industry. It will also provide opportunities for improving the working conditions of quality assurance testers by providing them more varied and skilled work opportunities.

Unique selling point

UC5 presents the opportunity to provide an underserved part of the game industry, quality assurance workers, with tools and workflows for automation, which will improve working conditions, bring games faster to the market, and de-risk game productions.

9.2.6 UC6: AI for Human Co-creation

→ Challenges

The integration of advanced AI tools for audio generation, and their interaction with non-expert users to create novel music, is a difficult challenge addressed from both a technical and an artistic perspective. Bringing to non-experts the possibility to use novel tools has the potential to enhance their capabilities to explore new artistic territories.

→ User needs

The target is an audience composed mainly by non-expert users of AI technology. These users however require the control of the full potential of these novel tools, mainly generative models, developed in a rather technical or even academic environment. As such, they are not robust and prepared for easy use.

→ Integration of novel AI functionalities

The goal is to integrate components developed for the generation and manipulation of audio tracks. The generation is performed by a collection of generative models, whose training can be performed under the supervision of the final user, or used after training performed elsewhere. Along the training, integration includes tools that may enhance the otherwise lengthy process. Additionally, generated audio files are manipulated to obtain information about their quality, and by extension of the overall generative process. This information can be used subsequently to further proceed in the content creation as a creative process.

→ Expected impact on media, society and democracy

Bringing powerful AI tools to a novel audience raises a number of questions difficult to address in an artistic environment, but we expect to expand the general use of these tools with easy to use integrations. With this, on one hand there is a positive impact on the community of artists and creators gaining access to novel creative tools, and on the other hand this serves as an excellent testbed and benchmarking platform for AI tools.

Unique selling point

UC6 develops an environment for the interaction of state of the art AI tools with advanced creators of music and explores the potential interaction between human creativity assisted by generative tools.

9.2.7 UC7: AI for Content Organisation and Content Moderation

→ Challenges

Implementation of powerful AI-powered image and video processing solutions requires investment in powerful GPU hardware, as well as machine learning specialists to finetune the solution for the media organisation's needs. Non-technical staff often have unrealistically high expectations that exceed the current capabilities of AI-based automated image and video organisation. A fast and accurate AI must be developed to overcome the hurdles to the efficient and cost-effective organisation of live streaming and videos that will be the prevailing media formats in the future.

→ User needs

Novel, creative methods for image/ video content (re)organisation are critical for the adaptation and survival of media companies in the era of user-generated content. Because of the high cost of professional content creation, there is a need for it to be re-purposed and re-used so it is economically viable. Investing in AI-enhanced tools for image/video re-organisation will make media companies better prepared for the future and the challenges it holds.

→ Integration of novel AI functionalities

To address user needs, the following AI functionalities are useful: tools for image/video analysis that will be able to understand context and properly understand abstract concepts; tools for fast and accurate identification of problematic content (propaganda, violence, infamous symbols, etc.); tools for real time analysis of multiple live streams and various video formats; tools optimised for fast and inexpensive analysis of big amounts of media content, including user generated content.

→ Expected impact on media, society and democracy

We aim at helping the media sector to smoothly and cost effectively transition and adopt AI automation by providing algorithms that consider the specificity of current digital content and thus minimise the efforts by a single media company to develop, train and adopt AI models. Media companies will be eased in organising, reusing and optimising their visual content while making sure it is safe for every member of the society and that the rule of law is followed.

Unique selling point

UC7 develops an automated visual content (re)organisation demonstrator integrating advanced deep learning techniques for tagging, categorisation, and facial recognition of media content to help media companies make sense of their exponential growing visual content and aid them in their efforts to showcase safe and politically correct visual content.

9.3 Intersection of AI4Media Use Cases & AI Research Themes

Table 1 illustrates the intersection of AI4Media Research Themes and the AI4Media Use Cases. Together, these lay the foundations for a new generation of AI methods and applications that will benefit the European media industry, society and democracy.

9.4 Expected outcomes

→ Real-world use case demonstrators based on tools and systems used by media organisations, which have been upgraded with novel AI functionalities.

→ **White papers for the AI4Media use cases** aiming to align AI research with media industry needs.

	UC1 AI against Disinformation	UC2 AI for News	UC3 AI in Vision	UC4 AI for Social Sciences	UC5 AI for Games	UC6 AI for Co-creation	UC7 AI for Content Organisation and Moderation
New learning paradigms							
Lifelong learning		●	●	●		●	●
Manifold learning	●	●		●	●	●	●
Transfer learning	●	●	●	●	●	●	●
Neural Architecture Search		●	●			●	
AI at the edge	●	●	●				
Deep quality diversity			●		●		
Learning to count	●	●	●				●
Quantum assisted RL			●				
Trustworthy AI							
Legal framework for trusted AI	●	●	●	●	●	●	●
AI robustness	●	●	●	●			●
AI explainability	●	●	●	●	●		●
AI fairness	●			●			
AI privacy	●			●			●
AI benchmarking	●		●	●			
Content-centered AI							
Media summarisation	●	●	●				
Media analysis	●		●	●			
Media content production		●	●		●	●	
Content representation, indexing and retrieval	●			●	●		●
Language analysis	●		●	●			
Music analysis				●	●		
Learning with scarce data			●	●			
Computationally demanding learning			●	●			●
Human-centered and Society-centered AI							
Content moderation							●
Manipulation and synthetic content detection	●	●		●			●
Content recommendation	●	●		●			
Political debate analysis	●		●	●			
Analysis of hyper-local news	●	●		●			
User perception of social media			●				
Real-life effects of private content sharing	●			●			

Table 1: AI4Media research themes and their connection to the seven use cases

10. AI education and AI skills development

Europe has no lack of talent. Through excellent Universities and higher education institutions but also new start-ups and SMEs, there is an abundance of young people with enormous potential. The ability to attract and retain top AI talent in Europe is crucial both to the effectiveness of the AI4Media research agenda and the success of European AI more broadly.

In support of this goal, the AI4Media Center of Excellence, aims to educate and improve skills of early career researchers and entrepreneurs that seek to work on media AI by introducing three initiatives:

- **The International AI Doctoral Academy**
- **The Junior Fellows Exchange Program**
- **The two AI4Media Open Calls**

10.1 International AI Doctoral Academy (AIDA)

→ Challenge

The academic standards of PhD acquisition in the AI domain vary among European universities. As a result, the set of typical scientific skills acquired during a PhD, as well as the transferable skills (e.g., communication, mobility, interpersonal), depend on the host institution's standards and practices. The aim of AIDA is to trigger a leading academic critical mass in Europe, so that excellence in AI research is achieved and industry focus is ensured.

→ Objective

The main objective of AIDA is to provide opportunities to PhD candidates/Postdocs/AI professionals hosted in AIDA academic/research/industrial institutions for scientific upskilling in the AI domain. The AIDA program will be a vehicle for providing access to top-quality academic material, in various formats, including academic courses (tutorials, short courses, semester courses, summer schools), free-access to thematically organised academic material (presentations, videos, review-papers), and lectures on hot AI topics **(AI Excellence Lecture series)**.

→ Opportunities & target groups

AIDA engages with academic professors, PhD students, junior/senior postdocs and AI professionals that are affiliated with AIDA members, or with well-recognized external professors in the wider AI domain. In AIDA terms, people from the above-mentioned target groups may act as AIDA Lecturers or AIDA students, on occasion.

AIDA Lecturers organise/upload academic offers, in the form of AIDA courses or other academic materials. AIDA offers a seal of Excellence to these materials with its curation processes, together with robust course dissemination services through AIDA channels that maximise course registration and impact. In case the lecturers are junior/senior postdocs, they may offer a course to AIDA to obtain teaching experience. On the other hand, AIDA students get the opportunity to attend quality AI courses delivered by distinguished AI lecturers worldwide, which increases their scientific skills. For each AIDA course they participate in, they may request and obtain their personal AIDA Certificate of Course Attendance, which strengthens their CV.

→ Expected impact on AI research and media industry

AIDA's expected long-term impact on AI research in Europe is the reduction of the variance in the typical academic skills obtained by AI PhD between the participating organisations, by the definition of commonly accepted academic curriculum and excellence PhD acquisition standards. The media industry and research organisations can benefit though AIDA by upskilling their own personnel and/or by hiring new personnel from the AIDA student pool.

Unique selling point

AIDA is a pan-European non-profit umbrella organisation that specializes in AI PhD education, uniting a leading academic mass in Europe to promote AI excellence and democratisation of education materials to PhD students among its members and beyond.

10.2 AI4Media Junior Fellows Exchange Program

→ Challenge

While AI is quickly reshaping the media industry landscape, media companies still face significant challenges in recruiting personnel that are skilled in AI. At the same time, AI researchers often lack insights on media industry needs. On top of that, mobility and collaboration between AI labs has yet to reach its full potential despite initiatives such as ELLIS. To establish Europe as a media AI powerhouse, it is necessary to promote exchange programs that will focus on AI research or applications for the media industry.

→ Objective

The Junior Fellows Exchange Program was launched in June 2021 to facilitate mobility of young researchers working on AI for media & society. It aims to develop new skills, improve diversity, increase visibility, and strengthen the impact of media AI through exchanges of researchers and media professionals between academia, research institutions, and media industry. Each exchange is expected to produce tangible open access outcomes (e.g. publications, software, datasets) on media AI while enabling the spread of expertise and skills across Europe and building strong collaborations between AI labs and/or media companies.

→ Opportunities & target groups

The Junior Fellows Exchange Program offers research exchange opportunities for PhD students, MS students, and early career postdocs but also senior researchers that want to improve their skills and knowledge in AI for the media and society. Each exchange involves one Fellow, one Sender institution, and one Host institution, with either the Sender or the Host being an AI4Media partner. Exchanges can be physical, virtual, or hybrid. The travel and living costs of physical exchanges are partly covered by the AI4Media mobility budget.

The topics of exchanges are relevant to the research directions use cases of AI4Media, extending existing AI4Media research and applications through the delivery of novel algorithms, software and datasets. Emphasis is put on promoting exchanges between media companies and AI labs to strengthen industry/research collaboration. These involve either AI researchers that seek to better understand media industry needs and gain practical experience on how to transform AI research into applications for the media or media professionals that seek to improve their AI skills and understanding.

→ Expected impact on AI research and media industry

The program is expected to contribute to the creation of a critical mass of early career researchers with a deeper understanding of media industry needs and significant experience in media AI research through their engagement with top media AI research labs and media companies in Europe. Both AI labs and media companies will benefit from the flow of novel ideas and the spread of media AI expertise and skills across Europe.

Unique selling point

The AI4Media Junior Fellows Exchange Program facilitates exchanges of early career researchers that want to improve their skills and knowledge in AI for the media and society by collaborating with top European AI researchers and media companies to conduct innovative research that considers industry needs.

10.3 AI4Media open calls

→ Challenge

AI is already a well-established but continuously evolving field, with applicability in various industries, including media and its many inherent areas. It is safe to assume that there are many promising ideas for AI research and solutions that often do not materialise due to the lack of opportunities and incentives that facilitate the growth of such ideas. Financial support to third party programmes (FSTP) is one such tool that can enable these ideas to grow.

→ Objective

The main objective of the AI4Media open calls is to engage entrepreneurs, companies (from start-ups to midcaps), academia and research organisations that develop and integrate applied research in the field of AI, to develop new and innovative research and applications for AI. The developed research and

solutions will contribute to the enrichment of the pool of research and technological tools to be made available – via AI4Media – on the AI-on-Demand platform.

→ Opportunities & target groups

The AI4Media project will run two open calls, funding a total of 20 projects. Each open call is divided into two tracks: Research and Application. Eligible entities include entrepreneurs, companies, academia and research organisations, which can apply to either track.

The projects funded in each track have submitted proposals addressing one of the many challenges defined for each open call. The challenges are prepared by the AI4Media consortium partners and are aligned with and complement the project's research and demonstration activities, as well as the AI4Media Roadmap on Media AI. The 20 projects funded will each receive up to €50,000 in equity-free funding, in addition to tailored training in the form of a bootcamp, business support to help promote project results, dedicated coaching support, and general promotion and visibility opportunities.

→ Expected impact on AI research and media industry

In addition to the direct impact that the funding opportunity has on the growth of the 20 participating entities, the integration of the developed research and solutions (e.g. related to music, recommendation systems, web traffic prediction, comment filtering, and others) in the AI-on-Demand platform will enable new research and applications to stem from those developed through the AI4Media open calls.

→ Open Call projects

Open Call 1 projects
Open Call 2 projects

Unique selling point

The AI4Media open calls will fund 20 projects led by industry or academia entities with up to €50,000 equity-free funding to develop new research in AI or innovative AI applications for the media sector.

11. Building the European Media AI Ecosystem

To advance its research agenda on media AI, AI4Media aims to establish a wide multi-disciplinary network of top AI research labs, social science labs, legal/ethics experts and media or media-related companies and start-ups across Europe that will closely collaborate to build excellence and create a European powerhouse on Media AI, aiming to benefit the media industry, the society and the economy.

The AI4Media consortium members, associate members, and organisations funded by the project's open calls already constitute a critical mass of more than 100 organisations from academia, research and industry, working together to produce excellent research and innovative applications for media AI, examining AI through a multi-lense including technology, industrial application, legal and ethical aspects, and socio-economic impact.

The concentration of knowledge and expertise in AI4Media makes the Network uniquely positioned to lead media AI research in Europe and to build collaborations on this topic. Besides supporting high-quality research, the project also supports engagement with industry to translate AI knowledge into practice; collaborations to spread AI expertise across research domains and countries; and training and education activities to develop the next generation of European AI talent.

→ Engaging industry to translate AI knowledge into practice

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.

As explained in section 8, AI4media aims to increase innovation in the media industry by translating mature research outcomes into industrial real-world applications through seven use cases. The use cases will be implemented in collaboration with well-known large European news media organisations such as DW in Germany, RAI in Italy and VRT in Belgium as well as rising SMEs in the media sector like modl.ai, a game development company in Denmark, and Imagga, company developing solutions for content management in Bulgaria.

In addition to the seven use cases, AI4Media will further support innovation acceleration activities through a programme of two open calls that will provide funding and support to SMEs and start-ups developing innovative AI applications for the media industry.

→ Collaborating to spread AI expertise across research domains

While AI4Media focuses on media AI, many of each research directions, including core ML research, analysis of multimedia content and trustworthy AI, have numerous applications in many other industry domains as well. To help build a European AI lighthouse and promote cross-domain and multi-discipline collaboration on AI, AI4Media has already identified complementary European initiatives with whom the Network could collaborate.

AI4EU, AI4Europe: To ensure that AI4Media's work is shared widely across the European innovation ecosystem, AI4Media closely collaborates with AI4EU and AI4Europe by:

- → Enriching the AI-on-Demand platform with AI4Media AI assets and integrating AI4Media components and pipelines in the AI4EU Experiments Marketplace.
- → Adopting, extending and operating the AI-Café, an online forum initially founded by the AI4EU project to gain insights into the European AI scene. Participants get the chance to share knowledge and experiences and meet stakeholders from various areas of AI research and application.
- → Participating in the AI4EU Technical Governance Board, through partner FhG-IAIS that also hosts and manages the AI4EU Experiments.

ICT-48 projects and other AI and Robotics NoEs: AI4Media has established important collaborations with the ICT-48 projects (TAILOR, ELISE, Humane AI, and VISION CSA) as well as with the new AI and

Robotics NoEs (ELSA, euROBIN, ELIAS, ENFIELD and dAIEDGE), aiming to create a European AI lighthouse. Complementing each other's expertise and know-how, the ICT-48 projects and new NoEs are well-positioned to establish Europe as an AI powerhouse, addressing the needs of different industries and approaching AI using a multi-disciplinary approach that considers research excellence, industrial needs, societal and economic impact, and legal-ethical issues. Collaboration with these projects projects spans different activities such as:

- AIDA foundation, management and operation (with ICT-48 projects).
- Organisation of cross-cutting events like Theme Development Workshops and AI Mellontology Symposia and AI Community Workshops.
- Development of a joint Strategic Research Agenda on AI, Data and Robotics.
- Development of an interactive map of the European AI ecosystem.
- Exchange of knowledge, expertise and best practices on a variety of topics from research to mobility and industrial collaboration.

EU projects on disinformation: AI4Media has established a collaboration cluster with EU projects that focus on disinformation, including veraAI[18], Titan[19], AI4Trust[20], AI4Debunk[21] and AI-CODE[22]. This collaboration has already resulted in joint events and publications. Such synergies can be leveraged to expand the reach of AI4Media's (trustworthy) AI technologies for the media sector with a focus on disinformation and to exchange research outcomes between the projects of this community.

European Digital Media Observatory (EDMO) and EDMO (multi)national Hubs[23]: EDMO brings together fact-checkers, media literacy experts, and academic researchers to understand and analyse disinformation, in collaboration with media organisations, online platforms and media literacy practitioners. EDMO's Hubs form a European multidisciplinary community aiming to actively detect, analyze and expose disinformation campaigns across Europe. Several AI4Media partners are members of EDMO and the Hubs. We aim to take advantage of these connections in order to promote AI4Media's outcomes and establish further collaborations in this area.

European Broadcasting Union (EBU)[24]: Through being a member in the EBU and/or specific working groups, some of the AI4Media media partners liaise with other public service media organisations in connection to selected AI & Media topics.

COPEAM[25]: To ensure its geographic reach in Southern Europe and the Mediterranean, AI4Media has forged a collaboration with COPEAM - an association of audiovisual players, including public service radio and TV broadcasters and professional and cultural associations of Balkans, North Africa and Middle East. The association is focused on training activities, knowledge sharing and awareness raising campaigns to which the AI4Media consortium will be contributing.

Better Images of AI (BIOAI)[26]: BIOAI advocates for the use of more representative and diverse imagery to be used when illustrating AI. One of its core activities is a repository of better images of AI that anyone can use, starting with a collection of inspirational images. The initiative explores what these new images might look like, and invites people from different creative, technical and other backgrounds to work together to create them. The initiative complements AI4Media's goal to demystify AI for the general public. AI4Media and BIOAI launched an open call to commission artists to create a catalogue of images that illustrate AI research ongoing in the project. These images are shared in BIOAI's repository.

→ Supporting the next generation of AI talent in Europe

AI4Media supports in practice the next generation of AI talent in Europe by founding and managing the International AI Doctoral Academy, a joint instrument to support world-level AI educational and training activities for PhD/postdoc researchers. In addition, through its flexible Junior Fellows Exchange Program, the Network offers mobility opportunities for young researchers and media professionals while its open call funding scheme supports SMEs, start-ups and research labs to develop innovative applications and research for the Media.

18 https://www.veraai.eu/
19 https://www.titanthinking.eu/
20 https://ai4trust.eu/
21 https://ai4debunk.eu/
22 https://aicode-project.eu/
23 https://edmo.eu/ and https://edmo.eu/edmo-hubs/
24 https://www.ebu.ch/home
25 http://www.copeam.org/
26 https://betterimagesofai.org/

→ Establishing the European AI infrastructure

To build the European AI Ecosystem, significant investments on AI infrastructure are required by the European Union and European industry. This should include the development of open AI platforms and resources similar to the AI-on-demand-platform (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; national AI research centers and European research networks 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.

12. Conclusion

This is the second version of the AI4Media Strategic Research Agenda which focuses on the use of AI technology in the service of the media, society and democracy. The first version of the Agenda presented four main research pillars focused on developing novel learning paradigms, trustworthy AI methods, multimedia content analysis and production tools, and human and society-centred AI applications for the media industry.

The second version introduces a new research pillar, namely Generative AI and LLMs, exploring recent developments and relevant challenges.

The Agenda also explores how the results of the aforementioned research can be integrated in seven media-related use cases, informed by emerging market opportunities and urgent media industry challenges.

The AI4Media Strategic Research Agenda also proposes a Media AI Observatory to monitor the legal and technological landscape as well as the impact of media AI while at the same introducing three initiatives (AI Doctoral Academy, Junior Fellows Exchange Program, and Open Calls) to advance AI education and skills development and to support entrepreneurship. Opportunities for collaborations that will help build the European Media AI Ecosystem are also discussed.

The Agenda should be seen as a snapshot of the current situation with regard to AI technology for the media industry and the research activities of the consortium. As AI and its applications for the media are advancing fast, new high-impact research areas and applications will be unveiled that will need to be explored. The Strategic Research Agenda will be updated to reflect these developments.

Table of Abbreviations and Acronyms

ABBREVIATION	MEANING
2D	Two-dimensional
3D	Three-dimensional
AI	Artificial Intelligence
AIDA	International AI Doctoral Academy
BIOAI	Better Images of AI
C2PA	Coalition for Content Provenance and Authenticity
COPEAM	Permanent Conference of Mediterranean Audiovisual Operators
CSN	Conditional Similarity Networks
DL	Deep Learning
DSA	Digital Services Act
DW	Deutsche Welle
EBU	European Broadcasting Union
EC	European Commission
EDMO	European Digital Media Observatory
EU	European Union
FhG-IAIS	Fraunhofer Institute for Intelligent Analysis and Information Systems
FSTP	Financial Support to Third Party
GAN	Generative Adversarial Network
GDPR	General Data Protection Regulation
GenAI	Generative AI
GPT	Generative Pre-trained Transformer
GPT-3	Generative Pre-trained Transformer 3
GPU	Graphics Processing Unit
ICT	Information Communication Technology
IoT	Internet of Things
ISCC	International Standard Content Code
LLM	Large Language Model
LMM	Large Multimodal Model
LtQ	Learning to Quantify
ML	Machine Learning
NAS	Neural Architecture Search
NeRF	Neural Radiance Field
NLP	Natural Language Processing
NN	Neural Network
PET	Privacy Enhancing Technologies
PPAI	Privacy-Preserving AI
PSM	Public Service Media
QA	Question Answering

Table of Abbreviations and Acronyms

ABBREVIATION	MEANING
QD	Quality-diversity
R&I	Research and Innovation
RAG	Retrieval-Augmented Generation
RAI	Radiotelevisione italiana S.p.A.
RIA	Research and Innovation Action
RL	Reinforcement Learning
RLHF	Reinforcement Learning from Human Feedback
SFL	Secure Federated Learning
SME	Small and Medium Enterprise
SRA	Strategic Research Agenda
SSH	Social Sciences and Humanities
TDM	Text and Data Mining
VLM	Vision Language Model
VR	Virtual Reality
VRT	De Vlaamse Radio- en Televisieomroeporganisatie nv
WWW	World Wide Web

Contributors

NAME	ORGANISATION
Yiannis Kompatsiaris	Centre for Research and Technology Hellas (CERTH)
Filareti Tsalakanidou	Centre for Research and Technology Hellas (CERTH)
Symeon Papadopoulos	Centre for Research and Technology Hellas (CERTH)
Vasileios Mezaris	Centre for Research and Technology Hellas (CERTH)
Nikos Giatsoglou	Centre for Research and Technology Hellas (CERTH)
Christoforos Papastergiopoulos	Centre for Research and Technology Hellas (CERTH)
Emine Ozge Yildirim	KU Leuven Centre for IT and IP Law (CiTiP)
Lidia Dutkiewicz	KU Leuven Centre for IT and IP Law (CiTiP)
Noémie Krack	KU Leuven Centre for IT and IP Law (CiTiP)
Nicu Sebe	University of Trento (UNITN)
Elisa Ricci	University of Trento (UNITN)
Cigdem Beyan	University of Trento (UNITN)
Lorenzo Seidenari	University of Florence (UNIFI)
Federico Pernici	University of Florence (UNIFI)
Ioannis Pitas	Aristotle University of Thessaloniki (AUTH)
Vasileios Mygdalis	Aristotle University of Thessaloniki (AUTH)
Ioannis Patras	Queen Mary University of London (QMUL)
Christos Tzelepis	Queen Mary University of London (QMUL)
Mihai Gabriel Constantin	University Politehnica of Bucharest (UPB)
Anna Schjøtt Hansen	University of Amsterdam (UvA)
Georgios Yannakakis	University of Malta (UM)
Antonios Liapis	University of Malta (UM)
Matthew Barthet	University of Malta (UM)
Roberto Gallotta	University of Malta (UM)
Henning Müller	University of Applied Sciences and Arts of Western Switzerland (HES-SO)
Riccardo Fratti	University of Applied Sciences and Arts of Western Switzerland (HES-SO)
Lorenzo Tronchin	University of Applied Sciences and Arts of Western Switzerland (HES-SO)
Adrian Popescu	French Alternative Energies and Atomic Energy Commission (CEA)
Julien Tourille	French Alternative Energies and Atomic Energy Commission (CEA)
Sven Becker	Fraunhofer Institute for Intelligent Analysis and Information Systems IAIS (FhG-IAIS)
Patrick Aichroth	Fraunhofer Institute for Digital Media Technology IDMT (FhG-IDMT)
Luca Cuccovillo	Fraunhofer Institute for Digital Media Technology IDMT (FhG-IDMT)
Thomas Köllmer	Fraunhofer Institute for Digital Media Technology IDMT (FhG-IDMT)
Frederic Precioso	Université Côte d'Azur (UCA)
Lucile Sassatelli	Université Côte d'Azur (UCA)
Fabrizio Sebastiani	National Research Council (CNR)
Andrea Esuli	National Research Council (CNR)
Nicola Messina	National Research Council (CNR)
Artur Garcia	Barcelona Supercomputing Center (BSC)
Adrian Tormos	Barcelona Supercomputing Center (BSC)
Daniel Gatica-Perez	Idiap Research Institute (IDIAP)
David Alonso del Barrio	Idiap Research Institute (IDIAP)
Victor Bros	Idiap Research Institute (IDIAP)

Contributors

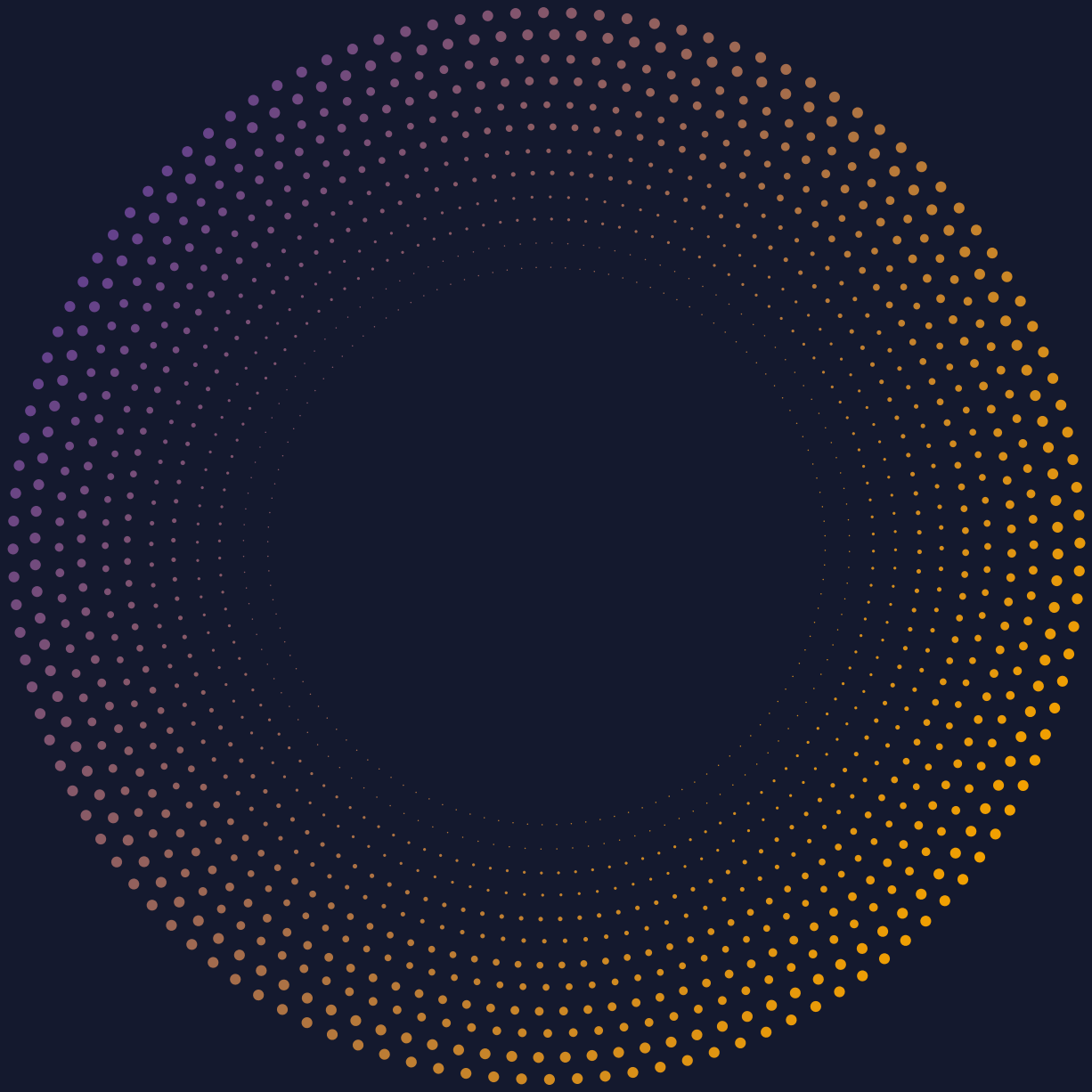
NAME	ORGANISATION
Hannes Fassold	Joanneum Research (JR)
Werner Bailer	Joanneum Research (JR)
Killian Levacher	IBM Research Ireland
Anisa Halimi	IBM Research Ireland
Danae Tsabouraki	Athens Technology Center (ATC)
Rasa Bočytė	Netherlands Institute for Sound & Vision (NISV)
Max Tiel	Netherlands Institute for Sound & Vision (NISV)
Johan Oomen	Netherlands Institute for Sound & Vision (NISV)
Birgit Gray	Deutsche Welle (DW)
Chaja Libot	Vlaamse Radio- en Televisieomroep (VRT)
Alberto Messina	Radiotelevisione Italiana (RAI)
Fulvio Negro	Radiotelevisione Italiana (RAI)
Samuel Almeida	F6S
Christoffer Holmgård	Modl.ai
Chris Georgiev	Imagga Technologies
Angel Spasov	Imagga Technologies
Candela Bravo	LOBA - Customer experience agency

AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

www.ai4media.eu

@ai4mediaproject



Our Consortium

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

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