AI4media

ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

D2.1

Overview & Analysis of the AI Policy Initiatives on EU level

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	Overview & Analysis of the AI Policy Initiatives on EU level
Deliverable number	D2.1
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	31 August 2021
Actual date of delivery	1 September 2021
Deliverable filename	AI4MEDIA_D2.1_final
Nature of deliverable	Report
Dissemination level	Public
Number of pages	138
Work Package	WP2
Task(s)	T2.1
Partner responsible	KUL
Author(s)	Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL), Noémie Krack (KUL), Lucile Sassatelli (UCA)
Editor	Aleksandra Kuczerawy (KUL)
EC Project Officer	Evangelia Markidou

Abstract	This Deliverable provides an analysis of the EU policy on AI and the recent Commission's legislative proposal on AI regulation. The aim is to provide a clear overview of existing and upcoming policy frameworks and an analysis of the ensuing principles and requirements.

Keywords	AI, Media, AI ethics, liability, IP law

Copyright

Contributors

NAME	ORGANISATION
Noémie Krack	KUL
Lidia Dutkiewicz	KUL
Emine Ozge Yildirim	KUL
Aleksandra Kuczerawy	KUL
Lucile Sassatelli	UCA
Filareti Tsalakanidou	CERTH

Peer Reviews

NAME	ORGANISATION
Rasa Bocyte	NISV
Georg Thallinger	JR

Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	12/02/2021	Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL), Noémie Krack (KUL)	Table of content, preliminary input
0.2	23/08/2021	Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL), Noémie Krack (KUL)	Pre-final version sent for internal review
0.3	27/08/2021	Rasa Bocyte (NISV), Georg Thallinger (JR)	Internal review
1.0	01/09/2021	Lidia Dutkiewicz (KUL), Emine Ozge Yildirim (KUL), Noémie Krack (KUL)	Final version ready for submission

Table of Abbreviations and Acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
AIA	Algorithmic Impact Assessment
ALTAI	Assessment list for trustworthy AI
Art.	Article
CAHAI	Ad Hoc Committee on Artificial Intelligence (Council of Europe)
CJEU	Court of Justice of the European Union
CoE	Council of Europe
Dir.	Directive
DIH	Digital Innovation Hub
EDIHs	European Digital Innovation Hubs
DGA	Data Governance Act
DMA	Digital Markets Act
DPIA	Data Protection Impact Assessment
DSA	Digital Services Act
EC	European Commission
EDPB	European Data Protection Board
EIC	European Innovation Council
EP	European Parliament
EPC	European Patent Convention
EPO	European Patent Office
EPRS	European Parliamentary Research Service
EU	European Union
EUR	Euro

Abbreviation	Meaning
GDPR	General Data Protection Regulation
GPAI	Global Partnership in AI
HLEG	High Level Expert Group
HITL	Human-in-the-loop
HIC	Human-in-command
HOTL	Human-on-the-loop
HPC	High Performance Computing
HRIA	Human Rights Impact Assessment
ICT	Information and communications technology
IEEE	Institute of Electrical and Electronics Engineers
IP	Intellectual Property
IPR	Intellectual Property Rights
ISO	International Organisation for Standardisation
IT	Information Technology
ITU	International Telecommuncations Union (United Nations)
IViR	University of Amsterdam Institute for Information Law
JIIP	The Joint Institute for Innovation Policy
KDT	Key Digital Technologies
NEPA	National Environmental Policy Act
OCDE	The Organisation for Economic Co-operation and Development
OSCE	The Organisation for Security and Co-operation in Europe
PLD	Product Liability Directive
POSITA	A Person Having Ordinary Skills in the Art
Q	Quarter of a year - period of 3 months
Rec.	Recital

Abbreviation	Meaning
R&I	Research and Innovation
SDGs	Sustainable Development Goals
SIA	Social Impact Assessment
SMEs	Small and Medium-sized Enterprises
STEM	Science, technology, engineering, and mathematics
T.	Task
TEFs	Testing and Experimentation Facilities
TFEU	Treaty on the Functioning of the European Union
TRIPS	The Agreement on Trade-Related Aspects of Intellectual Property Rights
UN	United Nations
VLOPs	Very Large Online Platforms
WP	Work Package

Index of Contents

Index of Tables

Index of Figures

1 Executive Summary

Deliverable D2.*1 "Overview & Analysis of the AI Policy Initiatives in EU level"* provides an overview of the EU policy initiatives on AI and the forthcoming Commission's legislative proposal on AI regulation. By doing so, the aim is mainly to provide a clear overview of existing and upcoming policy frameworks and an analysis of the ensuing principles and requirements to the AI4Media consortium.

Section 3 is arranged thematically. It presents and analyses EU initiatives on AI strategy, ethics, intellectual property rights, safety and liability, education, culture and audio-visual technology.

It starts with providing an overview of a broader context of AI regulation by outlining a selection of relevant initiatives from international bodies such as OSCE, CoE and OECD.

Section 3.2.1 analyses the overarching political AI initiatives taken by the European Commission and outlines the main features of the European Commission's key documents setting out the agenda for future AI policy and regulation. It includes the "Communication Artificial Intelligence for Europe", the "Coordinated Plan on Artificial Intelligence" and the "White Paper on Artificial Intelligence". These documents acknowledge the opportunities brought by AI as well as the risks to fundamental human rights and citizens' concerns. It is clear that a there is a need for a coordinated strategy on developing a common European approach to trustworthy AI. To that end, these documents call for the future regulatory framework which will address the risks associated with the AI technology.

Section 3.2.2 analyses policy documents related to AI Ethics. The EU's High-Level Expert Group (HLEG), "Ethics Guidelines for Trustworthy AI" together with its "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment" are the milestone documents in this field.

The AI HLEG Guidelines are centred around the concept of "trustworthy AI", which is based on three pillars: lawfulness, ethics and robustness. The HLEG Guidelines propose the four ethical principles which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner. The principles are then translated into a list of seven requirements to achieve Trustworthy AI. This deliverable provides consortium partners with a detailed assessment of these requirements: (i) human agency and oversight; (ii) technical robustness and safety; (iii) privacy and data governance; (iv) transparency; (v) diversity, non-discrimination and fairness; (vi) societal and environmental well-being; and (vii) accountability.

With the recognition of many benefits and potential risks AI technologies could bring, the European Commission and the European Parliament adopted different texts calling for harmonisation in order to avoid fragmentations of the Intellectual Property (IP) framework in the Union. To this end, Section 3.2.3 outlines objectives and recommendations of the European Parliament resolution on Intellectual Property Rights (IPR) for the development of AI

Technologies and the European Commission action plan on IP. It then provides a comprehensive analysis on the current state of art concerning IPR in the Union.

Artificial Intelligence in many of its aspects comes with promises and risks, the same goes for safety and liability. Section 3.2.4 deliberates on the European Commission studies and documents which outline the civil liability challenges raised by digital technologies and put forward recommendations on how to adapt the current legal framework on liability. It then turns to the European Parliament resolution and draft report with recommendations to the Commission on the adoption of a Civil liability regime for artificial intelligence.

Section 3.2.5 briefly outlines some other EU policy initiatives on AI, such as the EP resolutions on AI in criminal matters and in education, culture and the audio-visual sector, as well as EU projects related to AI technology (e.g. GAIA-X and quantum computing).

Section 4 provides an overview of EU regulatory initiatives in the field of AI. It starts with the analysis of the AI package published by the European Commission in April 2021. Part of the package is a proposal for a Regulation laying down harmonised rules on AI (Artificial Intelligence Act), which represents a key milestone on the way to a European approach to AI. This deliverable analyses the key features of the AI Act. A special attention is drawn to the prohibited AI practices and transparency obligations for chatbots, emotion recognition systems and 'deepfakes'. Subsection 4.2 presents the Digital Services Act Package. It provides an assessment of how the proposed rules to regulate digital services influence AI systems used for content moderation, displaying advertising, and recommender systems. Then, the Deliverable addresses the Digital Markets Act and the Data Governance Act, as well as the forthcoming Data Act.

Finally, the deliverable envisages that various anticipated and forthcoming EU policy and regulatory initiatives will have a profound impact both on research activities within the AI4Media project as well as on the commercial and non-commercial activities undertaken by AI4Media partners. Section 5 aims to anticipate this impact in four distinctive areas. First, the accessibility of social media data for researchers, fact-checkers and journalists is a major challenge. Recent regulatory initiatives, such as the Digital Services Act (see Section 4.2.), try to address this problem. Article 31 of the DSA proposal provides a specific provision on data access and scrutiny. The final scope of this provision will, undoubtedly, shape the way in which (vetted) researchers, journalists, and social activists will be able to access platforms' data. This is particularly relevant for the AI4Media WP6 ("*Human- and Society-centred AI*") activities such as opinion mining and automated extraction of public opinion from social media platforms such as Twitter that currently rely on the APIs. Second, academic research exception is only provided in a recital, and is not dealt with elsewhere in the proposed AI Act. It is not entirely clear whether this should be conceived as a general exception for research or a special exception only related to prohibited AI practices referred to in recital 16 (i.e., systems intended to distort human behaviour, whereby physical or psychological harms are likely to occur). Third, the scope of the AI Act is also unclear when it comes to transparency obligations applicable to bots, emotion

recognition systems and deepfakes (Art. 52 of the AI Act). It remains to be seen how the "emotion recognition system" definition and applicable transparency obligations for such systems change as the AI Act proposal follows the legislative process. In particular, 'sentiment analysis' and measuring and predicting user's affective response to multimedia content distributed on social media with the use of physiological signals (WP6) are likely to be considered as such. Fourth, some EU policy documents recommend the use of AI systems to detect IP infringements. The (il)legality of algorithmically filtered content should be subject to human review. The platform responsibility for third party infringing and/or illegal content will be particularly relevant in WP7 "*Integration with AI-On-Demand Platform*".

The detailed assessment of the impact of the EU regulatory initiatives will be conducted in the later stage of the project in D2.4 "*Pilot Policy Recommendations for the use of AI in the Media Sector*".

2 Introduction

In the last few years, Artificial Intelligence (AI) has gained prominence across individuals, businesses, academics and governments both at international and national level. Think-tanks, companies, and civil society have developed numerous toolkits, position papers and initiatives that focus on ethical principles for AI. A key player in the AI debate is the European Union (EU). As outlined in the Data Justice Lab working paper on 'European Artificial Intelligence Policy: Mapping the Institutional Landscape', while the popular interest in AI is a relatively new phenomenon, AI has been present in the European policy debate for quite some time.[1] However, recent years have brought an increased number of different policy documents which mention AI (Figure 1).

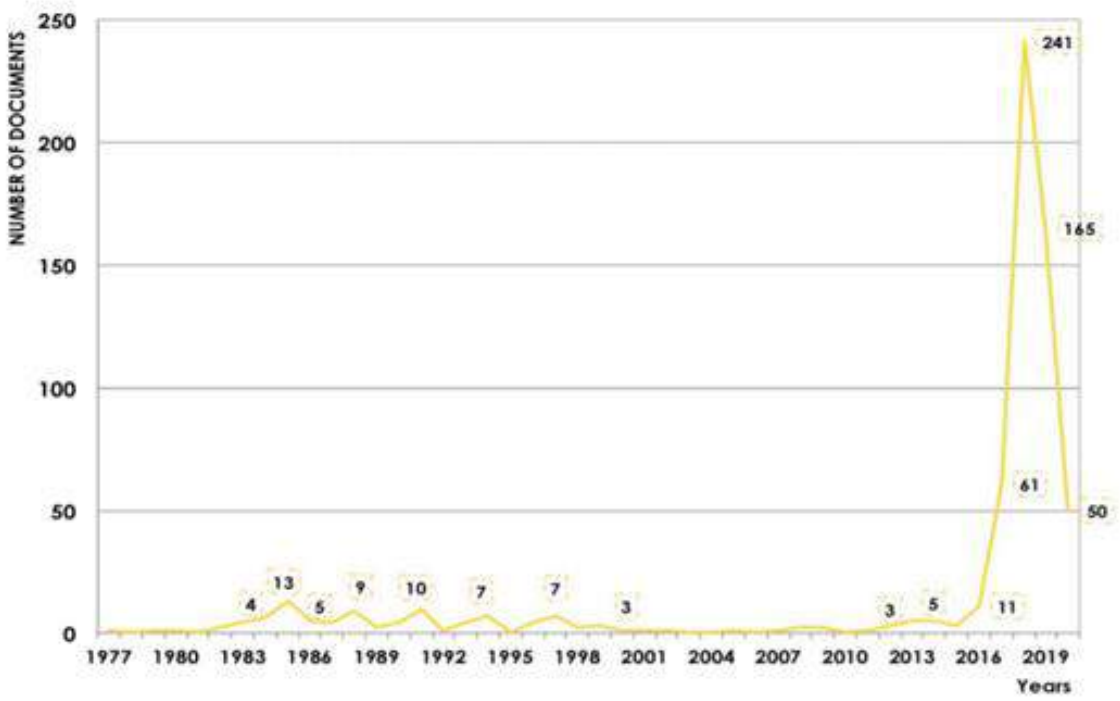

Figure 1: Mentions of AI in the EU legal and policy documents[2]

It is not surprising that the large number of developments in the EU in the area of the "AI policy initiatives" makes it very difficult for developers and researchers to monitor the ongoing debates and acquire a thorough analysis of the requirements. Keeping track of developments and trends in a rapidly evolving field of AI policy and regulation is a critical yet challenging endeavour.

[1] Niklas J, Dencik L, Working Paper on "European Artificial Intelligence Policy: Mapping the Institutional Landscape"', available at: https://datajusticeproject.net/wp-content/uploads/sites/30/2020/07/WP_AI-Policy-in-Europe.pdf.
[2] ibid.

It is important to note that there is already existing legislation at the EU level which continues to apply in relation to AI, although certain updates to that framework may be necessary to reflect the digital transformation and the use of AI. This EU legal framework consists of, inter alia:

- The EU framework for product safety and liability (e.g. The General Product Safety Directive (Directive 2001/95/EC), sector specific rules on machines, toys, cars, medical devices, etc.);
- The EU framework on fundamental rights (e.g. EU Charter, Race Equality Directive (Directive 2000/43/EC), Directive on equal treatment in employment and occupation (Directive 2000/78/EC), Directives on equal treatment between men and women in relation to employment and access to goods and services (Directive 2004/113/EC; Directive 2006/54/EC);
- The EU framework on fundamental rights and consumer protection (e.g. the Unfair Commercial Practices Directive (Directive 2005/29/EC), the Consumer Rights Directive (Directive 2011/83/EC);
- Data protection and privacy rules (e.g. General Data Protection Regulation).

This Deliverable provides an analysis of the EU policy on AI and the recent Commission's legislative proposal on AI regulation. We aim to provide the AI4Media consortium with an overview of existing and upcoming policy frameworks. We focus on the European Commission and other European institutions such as the European Parliament. However, where relevant, we refer to position papers, reports and studies produced by other stakeholders in the EU policy landscape.

The rest of this Deliverable is structured as follows:

- **Section 3**, AI Policy Initiatives in EU level, provides an analysis of the EU policy documents on AI:
    - Section 3.1, Context, takes a 'helicopter view' on various international and national AI policy initiatives.
    - Section 3.2, Existing AI Policy Initiatives in EU level, maps the main policy documents pertaining to AI:
        - ♣ Section 3.2.1, The overarching political AI initiatives, outlines key features of the European Commission's key documents setting out the agenda for future AI policy and regulation: Communication Artificial Intelligence for Europe, Coordinated Plan on Artificial Intelligence and White Paper on Artificial Intelligence.
        - ♣ Section 3.2.2, Ethics and trust AI initiatives, focuses on the Ethics Guidelines for Trustworthy Artificial Intelligence by the High-Level

Expert group on AI and provides a detailed analysis of its key principles and requirements.

- ♣ Section 3.2.3, Intellectual property rights AI initiatives, outlines key challenges and recommendations on AI and IP as identified by the European Parliament and the European Commission.
- ♣ Section 3.2.4, Safety and Liability initiatives, presents an overview of policy initiatives concerning the liability attached to AI systems. Various subjects of debate are presented such as: a one-size-fits-all framework, strict liability regime, reversal of the burden of proof, risk-based approach, maintenance of the traditional liability rules
- ♣ Section 3.2.5, Other policy initiatives, briefly touches upon some recent initiatives on the use of AI in criminal law and in the culture, education and audiovisual sector and technology (GAIA-X and quantum computing).

- **Section 4**, EU Regulatory initiatives in the field of AI, analyses forthcoming legislative proposals directly and indirectly pertaining to AI
  - Section 4.1, AI Package, analyses the key provisions of the Proposal for Artificial Intelligence Regulation (AI Act).
  - Section 4.2, The Digital Services Act Package, points out selected Digital Services Act provisions impacting the use of AI systems by intermediary services in the context of content moderation, online advertising and recommender systems. It also outlines key aspects of the Digital Markets Act.
  - Section 4.3, Data Governance Act, briefly outlines the EC initiative to enhance conditions for building common European data spaces.
  - Section 4.4, Data Act, sketches the European Commission plan to increase access to and further use of data.
- **Section 5**, The potential impact of the anticipated EU regulatory initiatives in the field of AI for AI4Media project, comprises of the likely consequences the forthcoming EU regulatory initiatives may have on the selected AI applications in the media sector;
- **Section 6**, Conclusions, ends with the final thoughts and next steps.

3 AI Policy Initiatives in EU level

3.1 Context – Plethora of Policy Initiatives on AI

Artificial Intelligence (AI) systems are reshaping our lives and constitute one of the major technological developments of our times. Hopes and fears are emerging in this relatively new field, and multiple voices and viewpoints take part in the AI debate. The question of creating a regulatory framework, ensuring the protection of users and conditions for innovation and technological progress of AI has grabbed public attention for the past few years. Having a transversal impact, AI systems have gained much scrutiny from a broad range of actors: international institutions, governments, stakeholders such as civil society, academia, private sector, etc. In this section we will sketch how diverse and multiple these initiatives are. This tangled landscape of documents makes it hard for non-experts to build knowledge and grasp landmark insights about AI policy initiatives.

3.1.1 International Institutions initiatives

This section outlines a selection of relevant initiatives from international institutions on AI.

The Organisation for Security and Co-operation in Europe - OSCE

The OSCE Representative on Freedom of the Media (RFoM) has set a specific focus on AI and freedom of expression and developed projects around it such as #SAIFE for spotlight initiatives on AI and freedom of expression.[3] In December 2020, it released a Policy Paper on freedom of the media and AI[4] and in April 2021, it published a policy paper on AI and freedom of expression in political competition and elections.[5]

OSCE will soon develop policy recommendations on the most effective ways to safeguard freedom of expression and media pluralism, when deploying advanced machine-learning technologies such as AI.

The Organisation for Economic Co-operation and Development – OECD

The OECD has been very active on the AI Policy scene. Firstly, the international institution has created the OECD.AI Observatory which provides information on AI from various resources, facilitates the dialogue between stakeholders while providing multidisciplinary, evidence-based policy analysis in the areas where AI has the biggest impact.[6] The website contains a lot of

[3] OSCE, "#SAIFE: Presentation of spotlight initiatives", https://www.osce.org/fom/ai-free-speech/spotlight-initiatives.
[4] OSCE Representative on Freedom of the Media, 'Policy paper on freedom of the media and artificial intelligence' (2020), 472488.pdf (osce.org).
[5] OSCE Representative on Freedom of the Media, 'Policy paper on AI and freedom of expression in political competition and elections' (2021), https://www.osce.org/representative-on-freedom-of-media/483638.
[6] OECD, 'The OECD Artificial Intelligence Policy Observatory', https://www.oecd.ai/.

interesting information on research, collaboration and policy initiatives on AI including sections about AI initiatives in different countries, statistics and trends about AI, specific policy areas focus and how AI impacts these aspects. It also includes the OECD AI principles which focus on how governments and other actors can shape a human-centric approach to trustworthy AI. They were adopted in May 2019, as an OECD legal instrument via the recommendation of the Council on Artificial Intelligence.[7] The document was ratified by 46 Countries.

The United Nations – UN

The United Nations opened a centre on Artificial Intelligence and Robotics which focuses on expertise on AI. The International Telecommunication Union (ITU), which is the UN's specialized agency for information and communication technologies and has also become key for assessing AI's impact. ITU has been organising for several years the AI for Good summits, which focus on how AI can accelerate the achievements of the UN Sustainable Development Goals.[8] ITU owns a journal where AI issues are regularly published and manages an AI repository identifying AI related projects, research initiatives, think-tanks and organizations that can accelerate progress towards the "17 UN Sustainable Development Goals (SDGs)".[9] ITU also publishes a report collecting the diverse UN activities on AI, the latest version available is the one of 2020.[10]

In April 2021, the United Nations released a Resource Guide on Artificial Intelligence Strategies, laying out existing resources on AI ethics, policies and strategies on national, regional and international level.[11]

Council of Europe - CoE

The Council of Europe is also a highly active actor on AI providing policy documents under the angle of fundamental rights. In September 2019, the Council of Ministers of the Council of Europe created an ad-hoc Committee on Artificial Intelligence (CAHAI). The CAHAI webpage contains a collection of relevant material from the Council of Europe[12], a collection of publications by scholars on AI and a data visualisation of AI initiatives categorised by the subject of or the entity responsible for the initiative.[13]

---

[7] OECD, 'Recommendation of the Council on Artificial Intelligence', (2019), OECD Legal Instruments.
[8] International Telecommunication Union (ITU), 'AI for Good', https://aiforgood.itu.int/.
[9] International Telecommunication Union (ITU), 'Journal on Future and Evolving Technologies', ITU Journal on Future and Evolving Technologies (ITU J-FET) and International Telecommunication Union (ITU), AI Repository, https://www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx.
[10] International Telecommunication Union (ITU), 'United Nations Activities on Artificial Intelligence 2020', https://www.itu.int/pub/S-GEN-UNACT-2020-1.
[11] United Nations, 'Resource Guide on Artificial Intelligence Strategies' (2021), Resource Guide on AI Strategies_April 2021_rev_0.pdf (un.org).
[12] Council of Europe, 'Council of Europe's Work in progress', (29.06.2021 last update), https://www.coe.int/en/web/artificial-intelligence/work-in-progress.
[13] Council of Europe, 'AI initiatives', https://www.coe.int/en/web/artificial-intelligence/national-initiatives.

An illustration of the policy actions undertaken by the Council of Europe is the recommendation on the human rights impacts of algorithmic systems.[14] Previously, a Declaration on the manipulative capabilities of algorithmic processes was also adopted.[15] In the beginning of 2021, the Council of Europe released guidelines on facial recognition.[16] Lately, a conference of ministers responsible for Media and Information Society held a conference on AI and the challenges and opportunities for media and democracy.[17] One of the topics covered was about the impacts of AI-powered technologies on freedom of expression, reflecting on a background paper written in 2020.[18] To conclude the conference, a declaration and resolutions were adopted.[19] They announce an engagement with all actors for AI tools used for the creation, moderation and distribution of online content in order to develop co-regulation or regulation - including through legally binding standards where appropriate - to ensure that freedom of expression is respected. This, including through tools such as natural language processing, robot-journalism and algorithmically prepared newsfeeds. Calls on private actors to pay attention to marginalised groups structurally excluded from receiving news and at risk of receiving a less diverse information offer were also made. In addition, a recently published guidance note on content moderation also deals with AI.[20]

Recently, attention focused on the long-term project of the CAHAI, which is to examine the feasibility and potential elements of a legal framework for AI. A document containing the global perspectives on the development of a legal framework on AI systems according to human rights, democracy and rule of law standards[21] was published in December 2020 along with the adopted

[14] Council of the Europe, 'Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems' (2020), https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154.

[15] Council of Europe, 'Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes' (2019), Result details (coe.int).

[16] Council of Europe, 'Guidelines on Facial Recognition' (28 January 2021), 1680a134f3 (coe.int).

[17] Council of Europe, 'Conference of Ministers responsible for Media and Information Society Artificial intelligence – Intelligent politics Challenges and opportunities for media and democracy' (2021), https://www.coe.int/en/web/freedom-expression/media2021nicosia.

[18] Council of Europe, 'Implications of AI-driven tools in the media for freedom of expression' (May 2020), 168097fa82 (coe.int).

[19] Council of Europe, 'Final Declaration of the Conference of Ministers responsible for Media and Information Society and resolutions on freedom of expression and digital technologies, on the safety of journalists, on the changing media and information environment, on the impacts of the COVID-19 pandemic on freedom of expression 10-11 June 2021', https://rm.coe.int/final-declaration-and-resolutions/1680a2c9ce.

[20] Council of Europe, 'Guidance Note - Content Moderation, Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation' (June 2021), 1680a2cc18 (coe.int).

[21] Council of Europe, 'Compilation of contributions, Towards Regulation of AI systems – Global Perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law' (December 2020), 1680a0c17a (coe.int)

Feasibility Study on a legal framework.[22] The study analyses the reasons why it is necessary today to have an adequate legal framework to protect human rights, democracy and the rule of law in light of the new challenges posed by AI systems. It also puts forward the different options and main elements of this framework: self-regulation, co-regulation, hard law. Various options are presented and explored. CAHAI then opened a multi-stakeholder consultation to gather the views of representative institutional actors on the key elements of this forthcoming framework and the form and scope it should take. The results of this consultation are not published yet. In June 2021, the Alan Turing Institute expanded on the ideas expressed in the Feasibility Study and supported readers with an accessible document.[23]

3.1.2 National initiatives

At the national level, policy initiatives flourished all over Europe and it is difficult or impossible to keep track of all of them. But as outlined above, some organisations provide mappings of the existing AI initiatives, including national ones such as the OECD AI Observatory or the Council of Europe. Another visible policy-making trend is the rapid emergence of regulatory instruments including self-regulation, sandboxes or legislative proposals. Many countries adopted national AI strategies and plans or undertook studies to assess and ensure the co-existence and respect of existing legislation by AI systems. Some guidelines or Codes of Practices were also adopted for the use of AI in certain sectors or containing general and horizontal principles.

Regarding legislative proposals, the initiatives found by the OECD.AI Policy Observatory are focusing primarily on regulating the use of automated vehicles in countries such as Germany, Austria, Denmark, Lithuania, Finland, and Spain. However, other initiatives focusing on ethics, national strategy and institutional organisation for dealing with AI can be mentioned for France, Malta, Denmark and Czech Republic. The assessment of these national efforts is beyond the scope of this deliverable. We refer to the work done by the OECD.AI Policy Observatory available at: https://www.oecd.ai/.

3.1.3 Stakeholder Initiatives

Stakeholders' initiatives are also countless, and the various sectors are extremely active in providing policy documents, reports, studies, surveys, etc. From the private sector, think-thanks, academia, civil society including NGO's, to professional association, technical community and trade unions, the number of their initiatives on AI has boomed in and since 2017 and a lot of attention was brought to the recent proposal released by the European Commission (EC) for an AI Act.

[22] Council of Europe, 'CAHAI Feasibility Study' (17 December 2020), 1680a0c6da (coe.int)
[23] Alan Turing Institute, 'A Primer Artificial Intelligence, Human Rights, Democracy and the Rule of law' (June 2021), 1680a2fd4a (coe.int)

3.1.4 Conclusion

To sum up and conclude this Section, AI has already caused a lot of ink to flow as demonstrated by the numerous, even countless policy initiatives on the matter. In this deliverable, we will focus on the policy initiatives at the EU level and beside this brief introduction and presentation will not dive deeper in non-EU documents. Every international institution and stakeholder wish to have their say in this moving field which can make it extremely hard for non-experts to access relevant information and make a choice within this myriad of policy documents and initiatives. A lot of attention has been brought to the development of AI systems but the matter having a horizontal impact on almost all aspects of our lives, the sector specific documents and the general documents are constituting a complex entanglement of guidelines, reports, studies and recommendation.

In the following subsection, a selection of existing AI policy initiatives at the EU level will be analysed.

3.2 Existing AI Policy Initiatives in EU level

In the last three years, there has been a variety of new publications, guidelines and political declarations from various EU bodies on AI. These documents provide a valuable insight into the future of AI regulation in the EU. This Section aims to provide an overview of all these policy initiatives.

3.2.1 The overarching political AI initiatives

3.2.1.1 Communication Artificial Intelligence for Europe

On April 25th, 2018, the EC issued a Communication on Artificial Intelligence for Europe.[24] The aim of the Communication is to embrace the idea that AI is transforming the world, the society, and the European industry. The EC notes that the way the EU approaches AI will define the world the EU citizens and other nationalities live in.

The EU's Position in the Globally Fierce Competition Concerning AI Landscape

According to the EC, while Europe is behind in private investments in AI,[25] countries like the United States and China, as well as large companies located in them, have been significantly investing in AI and are exploiting large amounts of data. Luckily, Europe has a strong industry

[24] European Commission, Communication Artificial Intelligence for Europe, April 2018. https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe.

[25] 10 Imperatives for Europe in the age of AI and automation, McKinsey, 2017. https://www.mckinsey.com/featured-insights/europe/ten-imperatives-for-europe-in-the-age-of-ai-and-automation.

concerning production of robots, manufacturing, healthcare, transport, and space technologies. It also plays an important role in the development and exploitation of platforms providing services to companies and organizations. It is also home to a world-leading research community, as well as innovative entrepreneurs and deep-tech start-ups.[26] Therefore, the Communication emphasizes that it is crucial for the EU to continue its efforts on creating an environment that stimulates investments and uses public funding to leverage private investments, while preserving and building on its assets. The EC adds that in order for the EU to be competitive, it needs to ensure the take-up of AI technology across its economy. In 2018, only a small fraction of European companies adopted digital technologies, despite that the benefits of adopting AI technologies are widely recognised.

The Commission notes that thanks to the EU's research and development framework with a specific focus on robotics launched in 2004, Europe has been gaining leadership in robotics. With this in mind, AI related research and innovation was included in the later phases of the Horizon 2020 programme, which is also funding the AI4Media project. Additionally, the Commission's major initiatives triggered the development of more efficient electronic components and systems, such as neuromorphic chips, world-class high-performance computers, quantum technologies, and technologies for the mapping of the human brain.[27]

However, in face of fierce global AI competition a solid European framework concerning AI technologies is crucial. Consequently, the EC is of the opinion that "the EU should have a coordinated approach to make the most of the opportunities offered by AI and to address the new challenges it brings." The end goal is to make the EU the leader in developing and using AI "for good and for all", as well as becoming the champion of an approach to AI that benefits people and the society as whole, building on European values and strengths. Therefore, according to the Communication, the EU can capitalize on:

(i) World-class researchers, labs, and start-ups;
(ii) The Digital Single Market, including the free flow of data in the EU; and
(iii) A wealth of industrial, research, and public sector data, which can be unlocked to feed AI systems.

[26] *The State of European Tech 2017*. https://2017.stateofeuropeantech.com/.
[27] European Commission, High Performance Computing Joint; Research and Innovation. https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing-joint-undertaking; https://ec.europa.eu/info/research-and-innovation_en.

The EU Initiative on AI

Finally, the Communication set out a European Initiative on AI,[28] which aims to boost the EU's technological and industrial capacity and AI uptake across the economy;[29] prepare for economic changes brought about by AI; and ensure an appropriate ethical and legal framework, based on the Union's values and in line with the Charter of Fundamental Rights of the EU.

Boosting the EU's technological and industrial capacity and AI update across the economy

The Communication emphasizes that the EU should be ahead of technological developments in AI and ensure they are swiftly taken up across its economy. Therefore, to achieve its aim of becoming the market leader in AI, the EC calls for the following efforts:

(i) Stepping up investments in AI related research and innovation framework;
(ii) Strengthening research and innovation from the lab to the market, including basic and industrial research;
(iii) Supporting and Strengthening AI research excellence centres across Europe, in addition to encouraging and facilitating their collaboration and networking;
(iv) Bringing AI to all small businesses and potential users by developing an "AI-on-demand platform" and creating a dedicated network of AI-centric Digital Innovation Hubs for better access;
(v) Supporting testing and experimentation infrastructure that are open to businesses of all sizes and from all regions;
(vi) Attracting a sufficient level of private investments in the AI transformation;
(vii) Making more (especially non-personal) data available for re-use while fully respecting personal data protection rules.[30]

Preparing for socioeconomic changes

The EC notes that AI could benefit society while the risks it is creating (or is expected to create) could be mitigated by taking the right action. For instance, utilizing AI technologies to enhance people's abilities has been gaining prominence. Moreover, new jobs and tasks will continue to emerge as a result of AI; meanwhile some other jobs and tasks will be replaced. Consequently, there are three main challenges the Communication wants to draw attention to. First, the society will need to be prepared as a whole. Second, the EU needs to focus efforts to help workers in jobs which are likely to be the most transformed or to disappear due to automation, robotics and AI. Third, the EU needs to train more specialists in AI, building on its long tradition of academic excellence, create the right environment for them to work in the EU and attract more talent from abroad. Hence, the Commission plans to provide up-skilling and training to

[28] European Council Meeting 14/17, October 2017. http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf.
[29] European Commission, The Ministerial Declaration on eGovernment – the Tallinn Declaration, October 2017. https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration).
[30] European Commission, Communication Artificial Intelligence for Europe, April 2018, op.cit.

foster digital skills and competence of all citizens, nurture talent, and encourage diversity and interdisciplinarity of AI trainings.

Ensuring an appropriate ethical and legal framework

In order to respect and enjoy the values set out in Article 2 of the Treaty on European Union and the EU Charter of Fundamental Rights, the Communication anticipates an appropriate legal and ethical framework. Some of the texts mentioned for the objectives set forth have already come to existence after the issuance of the Communication in 2018, i.e., the Ethics guidelines for trustworthy AI drafted by the High-Level Expert Group (2019)[31] and the Commission Report on safety and liability implications of AI, the Internet of Things and Robotics (2020).[32] Lastly, it goes without saying, the EC also highlights the importance of empowering individuals and consumers to make the most of AI.

Joining forces

Finally, the EC encourages Member States to engage in the coordinated plan on AI to share best practices, identify synergies, align actions, and fuel the emergence of AI start-ups while avoiding the fragmentation of the single market. It also foresees setting up a European AI Alliance to facilitate a broad multi-stakeholder platform to work on all aspects of AI and a systematic monitoring of AI development and uptake. Moreover, the EC wants to benchmark technical capabilities of AI components and systems and identify potential shifts in industrial value chains caused by AI, as well as societal and legal developments and the situation on the labour market.

3.2.1.2 Coordinated Plan on Artificial Intelligence

Delivering on its strategy on AI adopted in April 2018, on 8 December 2018, the Commission presented a coordinated plan prepared with Member States to foster the development and use of AI in Europe.[33] On 21 April 2021, the EC published the review of the Coordinated Plan (see Section 4.1.2).

The proposal of a coordinated plan built on the declaration of cooperation on AI launched in April 2018 at the Digital Day and signed by all Member States and Norway. It was endorsed by the European Council in June 2018. The Member States agreed to work together on the most important issues raised by AI, from ensuring Europe's competitiveness in the research and deployment of AI, to dealing with social, economic, ethical and legal questions. During the meetings taking place between June and November 2018, Member States, Norway, Switzerland

[31] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-AI.
[32] European Commission, Commission Report of 19 February 2020, on safety and liability implications of AI, the Internet of Things and Robotics. https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en.
[33] Communication Coordinated Plan on Artificial Intelligence.

and the Commission identified a series of common actions to increase investments, pool data - the raw material for AI -, foster talent and ensure trust. The result of this joint work in the Coordinated Plan. This plan builds on the idea that in order to ensure successful uptake of AI, coordination at European level is essential. It proposes joint actions for closer and more efficient cooperation between Member States, Norway, Switzerland and the Commission in four key areas:

Maximise investments

Under this heading, to complement national investments, the Commission committed to invest €1.5 billion by 2020. Other joint actions, to achieve these investment objectives include: (i) encouraging national AI strategies; (ii) a new European AI public-private partnership to foster collaboration between academia and industry in Europe and to define a common strategic research agenda on AI; (iii) new AI scale-up funding; (iv) developing and connecting world-leading centres for AI: European AI excellence centres.

Making more data available

The Plan comes with the realization that "making secure, robust quality data available for a broad range of users across borders is a cornerstone of European policy".[34] To make more data available and to facilitate sharing of data held by public and private sectors, the Commission commits to create a common European Data Space: a seamless digital area with the scale that will enable the development of new products and services based on data.

Nurture talent, skills and life-long learning

The Plan notes that EU countries face shortages of ICT professionals and lack AI-specialised higher education programmes. As a remedy, the Commission, together with Member States commits to supporting advanced degrees in AI and will support digital skills and lifelong learning.

Develop ethical and trustworthy AI

The Commission committed to firmly respect and anchor the "ethics by design" principle. The Plan also foresees that the Commission, taking into account the input from the Member States, will assess whether and to what extent the existing legislation is fit for purpose, taking into account the policy recommendations proposed by the AI High-Level Expert Group.

3.2.1.3 White Paper on Artificial Intelligence

On 19 February 2020, the EC released 3 key documents which set out its vision for the digital economy and its recommendations for digital policy making over the next five years: i) the

[34] ibid., p. 13.

European data strategy, ii) the Report on safety and liability implications of AI, the Internet of Things and Robotics, and iii) the White Paper on fostering trust and excellence in Artificial Intelligence.[35] The purpose of the White Paper on AI is to outline a strategy on developing a common European approach to trustworthy AI. The document analyses the strengths and weaknesses of the EU on AI but also the opportunities that AI can bring to the EU global market. The strategy outlined in the White Paper is based on European values and fundamental rights including human dignity and right to privacy but also on the sustainability dimension. The White Paper points out that AI will be key for meeting the European Green Deal Goals and underlines that the environmental impact of AI systems needs to be duly considered throughout their lifecycle. This includes not only the design, but also the storage of the data, the resources of usage and the waste management of the AI systems components.

To reach these goals, the AI strategy outlined in the White Paper details policy actions which will be undertaken to support the development and uptake of AI including investment increase, improve accessibility to data, and create a future regulatory framework which will address the risks associated with the AI technology. This last action point materialised with the release of the AI act proposal in April 2021 (see Section 4.1.3). The White Paper sets different policy and regulatory options and how to achieve the objectives. It suggests policy measures for an ecosystem of excellence and legal measures for an ecosystem of trust.

An ecosystem of excellence

The White Paper puts forward various measures organised around the following goals.

SHOWING LEADERSHIP IN AI

Under this heading, the EC wants to align the efforts at European, national and regional level as well as develop partnership between the private and public sector and academia. To solve the fragmented landscape of AI efforts and expertise, more synergies should be achieved notably with the creation of multiple European research centres of excellence and improve international cooperation in respect of fundamental rights and values.

INCREASING INVESTMENTS IN AI

The White Paper also foresees mobilising resources all along the value chain, starting in research and innovation to accelerate the development and uptake of AI. The objective is to attract over €20 billion of total investment in AI in the EU per year over the next decade. Also, a strong focus will be put on skills to fill competence shortages. This will be done by establishing networks of leading universities and higher education institutes to attract the best professors and scientists and offer world-leading masters programmes in AI, funded under Digital Europe. Access and use of AI by smaller organizations will also be promoted thanks to the Digital Innovation Hubs and the AI-on-demand platform.

[35] European Commission, White Paper on Artificial Intelligence – a European approach to excellence and trust, OJ COM(2020) 65 final, 19.02.2020, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:65:FIN.

ALIGN WITH THE DATA STRATEGY

The White paper underlines the importance of developing European Data Pools and ensure a good symbiosis with the European Data Strategy. Data are essential to fuel and train AI systems. Therefore, improving access to data, data infrastructures, and data management best practices are crucial to ensure the success of the European AI strategy.

The various measures to achieve the above goals are detailed below:

(i) Revision of the Coordinated Plan on AI, based on the results of the public consultation on the White Paper.
(ii) Creation of excellence and testing centres that can combine European, national and private investments, possibly including a new legal instrument, funded under Digital Europe and Horizon Europe.
(iii) Establish and support through the advanced skills pillar of the Digital Europe Programme networks of leading universities and higher education institutes to attract the best professors and scientists and offer world-leading masters programmes in AI.
(iv) Collaboration between Member States to ensure that at least one digital innovation hub per Member State has a high degree of specialisation on AI.
(v) Setting up of new public-private partnership in AI, data and robotics to combine efforts, and to ensure coordination of research and innovation in AI and a collaboration with Digital Innovation Hubs.
(vi) Promoting the adoption of AI by the public sector by initiating open and transparent sector dialogues giving priority to healthcare, rural administrations and public service operators through the 'Adopt AI programme' that will support public procurement of AI systems.

An ecosystem of trust

The White Paper acknowledges the risks and societal concerns coming along with the opportunities brought by AI and EU society concerns. On the one hand, citizens fear being left powerless in defending their rights and safety when facing the information asymmetries of algorithmic decision-making, while on the other hand companies are concerned by legal uncertainty and the impact this will have on AI uptake and their business activities. The positive opportunities of AI are clear, but there is a shortage of trust towards AI systems and risks of using AI for malicious purposes. In addition, Member States started regulating AI themselves with domestic legislation and it is important to avoid a market fragmentation due to a conflicting regulatory landscape. Harmonisation measures are needed to build trust and ensure the uptake of AI. These are the reasons why the White Paper suggests putting forward a new European legal framework.

On the material scope, the definition of AI will be strongly inspired and will build upon the definition of the AI High Level Expert Group.[36] The White paper underlines that the definition would need to be sufficiently flexible to accommodate technical progress while being precise enough to provide the necessary legal certainty. Regarding the territorial scope, the document suggests that requirements will be applicable to all relevant economic operators providing AI-enabled products or services in the EU, regardless of whether they are established in the EU or not. Fundamental rights will be a key component of the future framework including freedom of expression, non-discrimination, human dignity, and alike.

A new framework is needed as the current EU fundamental rights or products legislation are only covering certain types of situations and risks. There is also the changing nature of the AI's product needs to be considered in the new framework. In addition, some actors involved in the AI supply chain are not covered by the EU product safety legislation, for instance the third-party developers. The specific characteristics of many AI technologies render extremely hard to ensure compliance and therefore dedicated provisions on enforcement are necessary.

A risk-based approach is advanced to ensure a proportionate intervention from the EU legislator and avoid creating an excessive burden on SME's shoulders. The new framework will set up different categories accompanied with specific rules and requirements including a high-risk category. The White Paper insists on the fact that the determination of what is a high-risk AI application should be clear and easily understandable and applicable for all parties concerned.[37] Two cumulative conditions are put forward for the high-risk classification:

1. The AI application is in a sector with significant risks, such as healthcare, transport, energy and parts of the public sector, which should be specifically and exhaustively listed in the new legislation;
2. The AI application is used in such a manner that significant risks are likely to arise, such as producing legal or similarly significant effects for the rights of an individual or a company and posing risk of injury, death or significant material or immaterial damage.

Irrespective of the above, there might also be exceptional cases where the use of AI application should be always considered as high-risk such as recruitment processes and remote biometric identification.

MANDATORY LEGAL REQUIREMENTS

For determining the mandatory legal requirement which will be imposed on the relevant actors, the White Paper builds on the High Level Expert Group Guidelines for Trustworthy AI (for more

[36] "Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.", High Level Expert Group, a definition of AI, p.6.

[37] White paper on AI, op cit., p. 17.

information see Section 3.2.2). It also further suggested that different obligations should apply to different actors depending on who is best placed to address a potential challenge (the developer, the producer, the distributor, the importer, or the user of AI). The following key features were put forward (Table 1):

Training data	AI systems must be trained on data sets that are sufficiently broad and representative to avoid discrimination and that privacy and personal data are adequately protected during the use of AI-enabled products and services.
Data and record-keeping	Developers should keep accurate records regarding the data sets used to train and test the AI systems, data sets themselves in justified cases and documentation on the programming and training methodologies, processes and techniques used to build, test and validate the AI systems.
Information to be provided	For promoting the responsible use of AI, building trust and facilitating redress where needed, transparency is required. Clear information on the AI system's capabilities and limitations as well as requirements to inform users when they are interacting with an AI system and not a human being should be provided.
Robustness and accuracy	To ensure trustworthiness, robustness and accuracy is needed. AI systems need to be developed in a responsible manner and with an ex-ante due and proper consideration of the risks that they may generate. AI systems must adequately deal with errors or inconsistencies during all life cycle phases and ensure that the outcomes are reproducible.
Human oversight	To ensure a trustworthy and human centric approach, ensuring an appropriate involvement of human beings in AI systems is required. Human oversight will help AI systems to not undermine human autonomy or cause adverse effects. Human revision, monitoring or intervention afterwards and during the AI systems operation could be envisaged.
Specific requirements for certain particular AI applications, such as those used for purposes of remote biometric identification.	According to the current EU legislative framework (the GDPR and the Charter of Fundamental Rights), AI can only be used for remote biometric identification purposes where such use is duly justified, proportionate and subject to adequate safeguards (the Commission will launch a public consultation to determine the specific circumstances which would justify the use of remote biometric identification and determine common safeguards).

Table 1: White Paper suggestions for mandatory key requirements imposed on AI actors

Compliance and enforcement

For AI high-risk applications, prior conformity assessments like the ones already in place for the products on the EU market were put forward to assess the compliance with the mandatory requirements. These legal requirements will be enforced both by competent national and European authorities.

Voluntary labelling for no-high risk applications

Non-high-risk AI systems would not have to meet the legal requirements but could decide to make themselves subject, on a voluntary basis, either to those requirements or to a specific set of similar requirements established by a voluntary label. Once used, the requirements of the label would be binding.

Governance

A European governance structure on AI in the form of a framework for cooperation of national competent authorities is required to avoid responsibility fragmentation, and to increase the EU capacity for testing and certification of AI-enabled products and services. National competent authorities should play a key role in the future regulation's implementation and enforcement.

In the next section we will touch upon Ethics initiatives and the milestone document of the ethics guidelines for trustworthy AI, which influence the adoption of the White Paper and will shape the future of AI regulation.

3.2.2 Ethics and trust AI initiatives

3.2.2.1 The Ethics Guidelines for Trustworthy Artificial Intelligence

On 8 of April 2019, the EU's High-Level Expert Group (HLEG), a multi-stakeholder group of fifty-two experts, published the "Ethics Guidelines for Trustworthy AI".[38] The Guidelines are not legally binding. They do, however, pave the way for the future "AI regulation". The stakeholders are, moreover, encouraged to voluntarily opt to use these Guidelines.

In order to help operationalise the ethical requirements, the HLEG has also published the "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment"[39] and an online tool.[40]

[38] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).
[39] High-Level Expert Group on AI, 'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment' (European Commission 2019)'.
[40] Available at: https://futurium.ec.europa.eu/en/european-ai-alliance/pages/altai-assessment-list-trustworthy-artificial-intelligence.

Another useful resource was developed in the context of the AI4EU project by the research team of V. Dignum together with J.C. Nieves, A. Theodorou, and A. Aler Tubella: An abbreviated assessment list to support the Responsible Development and Use of AI.[41] As provided by authors, the abbreviated assessment framework is based on the Assessment List for Trustworthy AI (ALTAI). It is a self-assessment tool and can support organisations to perform a 'quick scan' of the AI-application they want to develop, procure, deploy, or use.

All questions are answered with a simple 3-point scale, indicating to what extent the criteria is met. The assessment levels are (from high to low):

(i) (2) the criteria are considered to be sufficiently and appropriately addressed, and evidence can be provided if necessary;
(ii) (1) the criteria have been considered but not fully satisfied;
(iii) (0) the criteria have not been addressed, or are considered not relevant to the application.

We consider this assessment list a useful resource, and we will refer to it later in the text. Obviously, as noted by authors, compliance with the list does not fully measure the 'quality' of the AI based system. It should also not be taken as a complete in-depth assessment or by any means as evidence of legal compliance.

3.2.2.1.1 Framework for Trustworthy AI

The AI HLEG Guidelines are centred around the concept of "trustworthy AI". Trustworthiness is defined "a prerequisite for people and societies to develop, deploy and use AI systems".[42] Without AI systems being trustworthy, unwanted consequences may arise and prevent the realisation of the social and economic benefits of AI.

According to the Guidelines, the three pillars of trustworthy AI are:

(i) **Lawfulness** – compliance with all applicable laws and regulations; those include: EU primary law (the Treaties of the EU and its Charter of Fundamental Rights), EU secondary law (such as the General Data Protection Regulation, the Product Liability Directive, the Regulation on the Free Flow of Non-Personal Data, anti-discrimination Directives, consumer law and Safety and Health at Work Directives), the UN Human Rights treaties and the Council of Europe conventions (such as the European Convention on Human Rights), and EU Member State laws. Various sector-specific rules that apply to particular AI applications should also be taken into consideration;

[41] Dignum V. and others, 'An Abbreviated Assessment List to Support the Responsible Development and Use of AI' 11.
[42] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019. (n 31).

(ii) **Ethics** – respect for ethical principles and values; and

(iii) **Robustness** - both from a technical and social perspective, individuals and society must also be confident that AI systems will not cause any unintentional harm. AI should perform in a safe, secure and reliable manner, and safeguards prevent any unintended adverse impacts of AI applications should be put in place.[43]

Moreover, according to the HLEG, the foundations of Trustworthy AI are the fundamental rights enshrined in the EU Treaties, the EU Charter and international human rights law:

(i) Respect for human dignity, which should never be diminished, compromised or repressed by others – nor by new technologies like AI systems;[44]
(ii) Freedom of the individual, including freedom from (in)direct illegitimate coercion, unjustified surveillance, deception and unfair manipulation;
(iii) Respect for democracy, justice and the rule of law. In particular, AI systems must not undermine democratic processes, human deliberation or democratic voting systems.[45]
(iv) Equality, non-discrimination and solidarity - including the rights of persons at risk of exclusion. AI systems should generate unfairly biased outputs, meaning that the data used to train AI systems should be as inclusive as possible, representing different population groups.[46]
(v) Citizens' rights, In particular, AI systems should not negatively impact citizens' rights, including the right to vote.

Importantly, many of these rights are, to some extent, legally enforceable in the EU, so that compliance with their terms is legally obligatory. Besides legally enforceable rules, ethical guidelines can help to "identify what we should do rather than what we (currently) can do with technology".[47]

The schematic overview of the AI HLEG Framework for Trustworthy AI is illustrated below (Figure 2).

[43] ibid. 7.
[44] ibid. 10.
[45] ibid. 11.
[46] ibid. 11.
[47] ibid. 10.

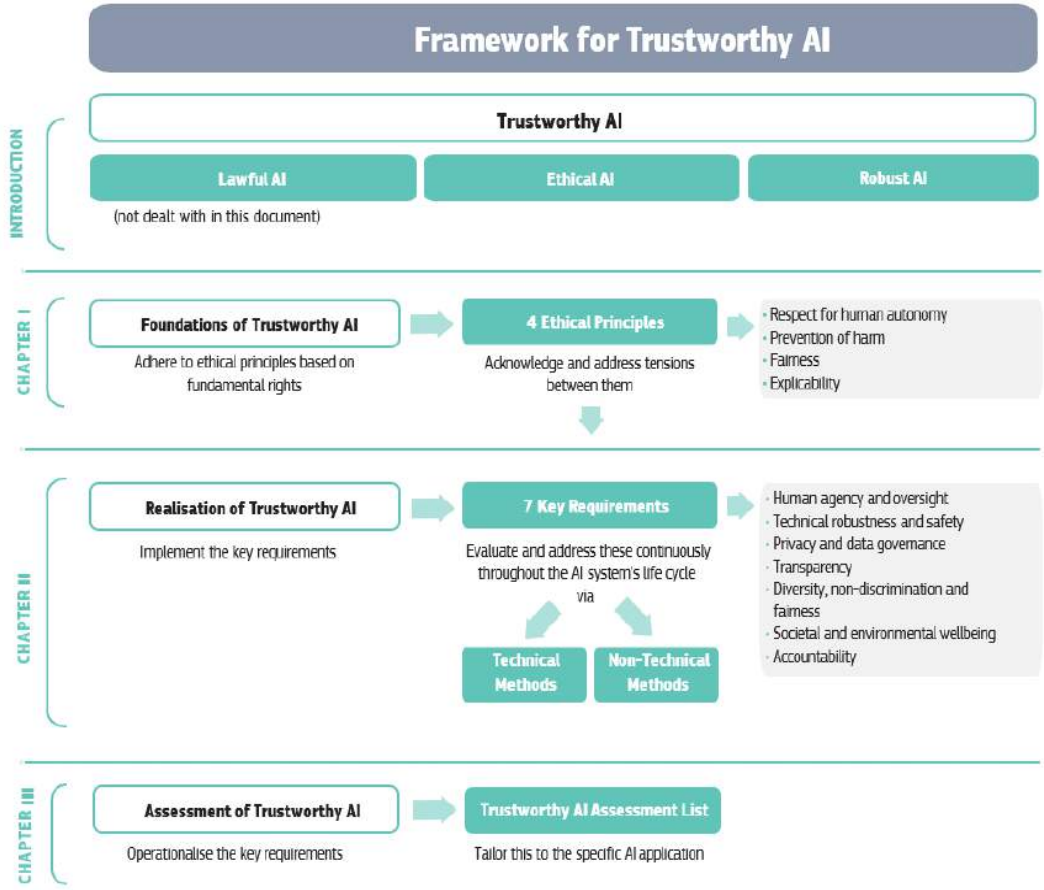

Figure 2: Framework for Trustworthy AI.[48]

3.2.2.1.2 The principles of trustworthy AI

The HLEG Guidelines propose the following four ethical principles ("imperatives") "which must be respected in order to ensure that AI systems are developed, deployed and used in a trustworthy manner"[49]: (i) the principle of respect for human autonomy; (ii) the principle of prevention of harm; (iii) fairness; and (iv) the principle of explicability.

The principle of respect for human autonomy

In moral theory, Kant found that "autonomous people are considered as being ends in themselves. In that, they have the capacity to determine their own destiny, and such must be respected."[50] Respect for human autonomy, as the AI HLEG emphasizes, means ensuring that

[48] ibid. 8.
[49] ibid. 11.
[50] Banham (2007) Kant's Moral Theory, British Journal for the History of Philosophy, 15:3, 581-593, DOI: 10.1080/09608780701445136

humans interacting with AI systems must be able to keep full and effective self-determination over themselves. Therefore, AI systems should follow human-centric design principles, augmenting, complementing, and empowering human cognitive, social, and cultural skills. Though the Guidelines do not explicitly state, human autonomy is closely linked with the right to integrity of the person, whose protection is enshrined in Art. 3 of the EU Charter of Fundamental Rights.[51] Thus, not only bodily integrity is provided safeguards under the Charter, but also mental integrity enjoys the same level of protection. Therefore, unjustifiably coercing, deceiving, manipulating, conditioning, or herding humans goes against the principle of respect for human autonomy. In the AI development context, this principle translates into ensuring human oversight over work processes in AI systems, to leave meaningful opportunity for human choice, as well as ensuring redress mechanisms to challenge harms to liberties caused by such systems.[52] As Floridi and others emphasize, there needs to be a balance between decision-making power that is freely given by the user to the autonomous systems and when this option is taken away or undermined by the system.[53] In other words, users should be informed actors and have control over their decisions when interacting with AI.[54] Consequently, AI should be used to empower, strengthen and, respect individual liberties, outlined in the EU Charter, Universal Declaration of Human Rights, the European Convention of Human Rights, etc., rather than curtailing or infringing upon them.[55] Lastly, the AI HLEG draws attention to the power of AI systems fundamentally changing the work sphere. Therefore, these systems should support humans in the working environment, and aim for the creation of meaningful work. This is also related to the principle of fairness, discussed below, in the sense that there should be effective, non-discriminatory, and fair ways to retrain, retool, and respect human workforce when AI replaces many human jobs in the future.[56] Furthermore, if AI is used within the judicial system, accountability should still lie with the human user to avoid unjust and unfair outcomes.[57]

[51] https://fra.europa.eu/en/eu-charter/article/3-right-integrity-person

[52] ICO (2017), "Big data, artificial intelligence, machine learning and data protection", 2017, <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.

[53] Floridi and others, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 Minds and Machines 689.

[54] Council of Europe, "European ethical charter on the use of artificial intelligence in judicial systems and their environment,'' 2019. <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>.

[55] International Conference of Data Protection and Privacy Commissioners (ICDPPC) (2018), "Declaration on ethics and data protection in artificial intelligence." <http://globalprivacyassembly.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf>.

[56] COMEST/UNESCO, "Report of COMEST on robotics ethics", 2017. <https://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20COMEST%20ON%20ROBOTICS%20ETHICS%2014.09.17.pdf>.

[57] Rathenau Institute "Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality", 2017. <https://www.rathenau.nl/en/digitale-samenleving/human-rights-robot-age>.

The principle of prevention of harm

The AI HLEG emphasizes that the principle of prevention of harm is one of the crucial ethical principles during and after the development and design of AI systems. Thus, these systems should not cause or exacerbate harm to human beings, nor should they otherwise adversely affect human beings. The principle also entails the consideration of the natural environment and all living beings. Conclusively, AI systems should be designed with the intent of not doing foreseeable harm,[58] by ensuring that they are technically robust and operate in a safe and secure environment. Prevention of harm, similar to the principle of respect for human autonomy, is intimately linked to the right to integrity of the person, as well as the protection of human dignity. Therefore, it is crucial for developers to ensure that AI does not infringe on human rights or cause "bodily injury or severe emotional distress to any person"[59] by assessing their technologies' safety, testing their algorithms to determine that no harm results from them, and implementing algorithmic accountability standards for any foreseeable negative impacts.[60] Furthermore, developers and organisations using AI should receive and follow the advice of legal authorities and research ethics boards concerning personal and otherwise data collection and processing.

In addition to the aforementioned main components of the principle, the HLEG highlights that vulnerable persons should receive greater attention and be included in development, deployment, and use of AI systems. Hence, particular attention must also be paid to situations where AI systems can cause or exacerbate adverse impacts due to asymmetries of power or information (i.e., power relations between employers and employers, businesses and consumers, governments and citizens, or vulnerable persons and persons not deemed vulnerable.) The power asymmetries also create the need to allow external auditors to conduct examinations and report negative impacts of the AI without fear of harm or threat by the AI organisations. Moreover, along with ensuring not to encumber external audits, AI organisations should also provide protection for whistle-blowers within the organisation to allow for effective and legitimate reporting of harms. Lastly, according to civil society experts, "AI systems should allow for human interruption, or their shutdown, when there is potential harm."[61] Thus, AI should "fail gracefully" if the reliability, safety, and internal robustness of the systems cannot be

[58] Personal Data Protection Commission Singapore, "Discussion paper on AI and personal data — fostering responsible development and adoption of AI", 2018, <https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/Discussion-Paper-on-AI-and-PD---050618.pdf>.
[59] Icelandic Institute for Intelligent Machines, "Ethics policy", 2015. <https://www.iiim.is/ethics-policy/>.
[60] Algo.Rules , "Rules for the design of algorithmic systems", 2019, <https://algorules.org/en/home>.
[61] Internet Society, "Artificial intelligence and machine learning: policy paper", 2017. <https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/>.

ensured.[62]

Fairness

The principle of fairness is probably the most crucial of all the principles laid out in the HLEG Guidelines, after the principle of respect for human autonomy. This is due to the reason that in order to guarantee prevention of harm, respect for human autonomy, and allowing the realisation of trustworthy AI in practice, fairness should be ensured first and foremost. It is important to note that, according to some civil society experts, "all automated systems make decisions that reflect bias and discrimination, but such decisions should not be normatively unfair."[63] However, there is no simple answer or definitions to what is deemed unfair, nor the fairness principle could be reduced into an assessment of objective outcomes without evaluating normative consequences pre-existing or amplified by an AI system.[64]

While the AI HLEG is aware that there are many different interpretations of fairness, they emphasize that the development, deployment, and use of AI systems must be fair. To comply with this principle, AI design should be "fit for purposes, identity impacts on different aspects of society and should be designed to promote human welfare, rather than endanger it."[65] Thus, in order to make the principle's interpretation more explainable, the AI HLEG lays out a two-dimensional definition for the principle of fairness: 1) substantive dimension, and 2) procedural dimension.

First, the substantive dimension refers to:

(i) Ensuring equal and just distribution of both benefits and costs;
(ii) Ensuring that individuals and groups are free from unfair bias, discrimination, and stigmatisation;
(iii) Ensuring equal opportunity in terms of access to education, goods, services, and technology. AI should be accessible to those that are often "socially disadvantaged"[66], such as disabled people;
(iv) Preventing deception of people or unjustifiable deterioration of their freedom of choice;
(v) Respecting the principle of proportionality between means and ends;
(vi) Considering carefully how to balance competing interests and objectives.

[62] IEEE, "Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems", Version 1, 2019. <https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf>.
[63] The Public Voice, AI Universal Guidelines, 2018. <https://thepublicvoice.org/ai-universal-guidelines/>.
[64] ibid.
[65] ibid., p. 54.
[66] Sage, "The ethics of code: Developing AI for business with five core principles", 2017. <https://www.sage.com/~/media/group/files/business-builders/business-builders-ethics-of-code.pdf>.

Second, the procedural dimension of fairness, in other words process fairness, in decision-making[67] entails the ability to challenge, contest, seek effective redress and remedies, and hold entities, AI systems, and human operators responsible accountable.[68] Therefore, procedural fairness provides for a fair chance to receive information, respond, and dispute for the affected party. Thus, it is closely related to the principle of explicability and the requirement of transparency, which are discussed below.

Consequently, there should be steps in place to ensure that data being used by AI does not contain errors, inaccuracies, and historical or unjust bias.[69] Ensuring the abovementioned is especially important when an AI system's decision-making process might potentially result in unfair outcomes. Therefore, inclusion, non-bias, equality, consistency, equal access, and equity approaches should be given a special prominence during the life-cycle of AI systems, as well as the mechanisms of the procedural dimension.

The principle of explicability

The AI HLEG sees explicability as a key principle for building and maintaining users' trust in AI systems. As noted by Robbins, "it is rare to see large numbers of ethicists, practitioners, journalists, and policy-makers agree on something that should guide the development of a technology. Yet, with the principle requiring that AI be explicable, we have exactly that." [70] He notes that not only the AI HLEG, but also Microsoft, Google, the World Economic Forum and academics[71], all include a principle for AI that falls under the umbrella of 'explicability'. However, what 'explicability' means varies. According to AI HLEG's Guidelines, explicability encompasses both transparency and explainability. In HLEG's own words "[explicability] means that processes need to be transparent, the capabilities and purpose of AI systems openly communicated, and decisions – to the extent possible – explainable to those directly and indirectly affected."[72] Moreover, in HLEG's Guidelines, explicability is not a means in itself. It should provide such information about the AI system which would enable a decision to be duly contested.

Nevertheless, in "black box society" as famously put by Frank Pasquale[73] an explanation as to why a model has generated a particular output or decision is not always possible. In these cases,

[67] Grgic-Hlaca, Nina, M. B. Zafar, K. Gummadi and Adrian Weller. "Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning", 2018.
[68] Ryan M, Stahl BC, Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, Journal of Information, Communication and Ethics in Society, 2020.
[69] ibid., p. 60.
[70] Scott R, 'A Misdirected Principle with a Catch: Explicability for AI' (2019) 29 Minds and Machines 495.
[71] Floridi and others (n 53).
[72] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).
[73] Pasquale F., *The Black Box Society: The Secret Algorithms That Control Money and Information* (Harvard University Press 2015).

the AI HLEG Guidelines note, other explicability measures such as traceability, auditability and transparent communication on system capabilities may be required.

Renda finds the principle of explicability of AI systems "perhaps the most controversial imperative".[74] He argues that in certain circumstances "invoking the full explicability of AI systems and decisions could jeopardize the use of AI techniques".[75] Along the same vein, Robbins argues that principles requiring AI to be explicable are misguided and the property of 'requiring explicability' is incorrectly applied to AI.[76] The real object which requires explicability is the result of the process—not the process itself. Instead of trying to achieve what (often) seems impossible, namely having powerful AI that can explain its decisions, we should be deciding which decisions require explanations.[77] Luciano Floridi proposes to develop a framework to enhance the explicability of those AI systems that make "socially significant decisions".[78] To this end, the HLEG clarified that "the degree to which explicability is needed is highly dependent on the context and the severity of the consequences if that output is erroneous or otherwise inaccurate".[79]

3.2.2.1.3 The requirements of trustworthy AI

In chapter II of the guidelines, the abovementioned principles are translated into a list of seven requirements to achieve Trustworthy AI:

Human Agency and Oversight

The first ethical requirement of the Guidelines concerns that AI systems shall respect human agency and oversight. Thus, AI systems should support human autonomy and decision-making, in accordance with the principle of respect for human autonomy prescribed by the Guidelines. This requirement sets forth that AI systems shall act as enablers to a democratic, flourishing, and equitable society by supporting the user's agency, upholding fundamental rights, and allowing for human oversight. In other words, an AI system shall not unjustifiably subordinate, coerce, deceive, manipulate, condition, and herd humans, while humans interacting with the system must be ensured to have full and effective self-determination over themselves and partake in democratic processes.

---

[74] Renda A., 'Europe: Toward a Policy Framework for Trustworthy AI' in Markus D Dubber, Frank Pasquale and Sunit Das (eds), Andrea Renda, *The Oxford Handbook of Ethics of AI* (Oxford University Press 2020) <https://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190067397.001.0001/oxfordhb-9780190067397-e-41> accessed 12 October 2020.

[75] ibid.

[76] Robbins (n 70).

[77] ibid.

[78] Floridi and others (n 53).

[79] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).

The EU is founded upon the fundamental rights that are directed towards ensuring respect for the freedom and autonomy of human beings. In order to comprehend the risk-benefit analysis of such systems on these rights, it is crucial to keep in mind that AI systems, like many other technologies, have the capacity to equally impact fundamental rights both negatively and positively. Therefore, to mitigate the situation where risk of negative impact exists:

(i) A fundamental rights impact assessment should be undertaken, prior to the development of such system;
(ii) This assessment should include an evaluation of whether those risks can be reduced or justified as necessary in a democratic society in order to respect the rights and freedoms of others;
(iii) Other mechanisms should also be put into place to receive external feedback regarding AI systems that potentially infringe on human rights.

Finally, the Guidelines evaluate this ethical requirement under two sub-sections: (1) Human Agency and (2) Human Oversight.

Human Agency

AI systems can have an effect on human behaviour in the widest sense, as the effect of such systems could be aimed at guiding, influencing, or supporting humans in decision-making processes (i.e., algorithmic decision support systems, risk analysis/prediction systems, recommender systems, predictive policing, financial risk analysis, and alike). Additionally, the effect on human perception and expectation when confronted with AI systems that 'act' like humans, as well as such systems' effect on human affection, trust, and (in)dependence should be considered as one of the major themes of this sub-requirement. Thus, the Guidelines suggest that the overall principle of user autonomy must be central to any AI system's functionality. Additionally, the right not to be subject to a decision based solely on automated processing when this produces legal effects on users should be one of the key components of the application of this sub-requirement. The following should be ensured to achieve the objectives mentioned above: (i) allowing users to be able to make informed autonomous decisions regarding AI systems; (ii) giving users the knowledge and tools to comprehend and interact with AI systems to a satisfactory degree; and (iii) enabling them to reasonably self-assess or challenge the system. Consequently, AI systems should support individuals in making better, more informed decisions in accordance with their goals. In other words, they should leave meaningful opportunities for human choice.

Human Oversight

Referring to the principle of respect for human autonomy, this sub-requirement ensures that an AI system does not undermine human autonomy or causes other adverse effects to human rational control. This objective could be achieved through governance mechanisms such as the following approaches:

1. Human-in-the-loop (HITL) means the capability for human intervention in every decision cycle of the system, which in many cases is neither possible nor desirable.
2. Human-on-the-loop (HOTL) means the capability for human intervention during the design cycle of the system and monitoring the system's operation
3. Human-in-command (HIC) means the capability to oversee the overall activity of the AI system (including its broader economic, societal, legal and ethical impact) and the ability to decide when and how to use the system in any particular situation. This can include the decision not to use an AI system in a particular situation, to establish levels of human discretion during the use of the system, or to ensure the ability to override a decision made by a system.

These oversight mechanisms can be required in varying degrees to support other safety and control measures, depending on the AI system's application area and potential risk. Additionally, it must be ensured that public enforcers have the ability to exercise oversight in line with their mandate. Finally, and most especially, the less oversight a human can exercise over an AI system, the more extensive testing and stricter governance must be required.

To conclude, despite seemingly providing strong ethical safeguards, this requirement has also received many criticisms from stakeholders, including academics, industry players, and civil society organizations. First, the requirement's emphasis on putting in mechanisms to receive external feedback regarding AI systems that potentially infringe fundamental rights does not clarify who should provide this feedback and whether the feedback is binding and to what extent. This ambiguous language could give rise to not adequately protecting human rights in AI systems' extensive information asymmetry.[80] Second, it is not clear how the values articulated by the HLEG would be balanced against each other. The requirement of diversity, fairness, and non-discrimination, which is discussed below, could, for example, push recommender systems to try to reach a broader audience or promote a more diverse content to advance values of non-discrimination and diversity. However, this could also be seen as interfering with human autonomy by nudging people toward content that they would otherwise not choose to engage with.[81] Third, in human autonomy and its protection by human agency and human oversight requirement context, many fundamental rights concerning mental integrity, such as freedom of expression and freedom of thought and conscience, have not been taken into account enough. It seems like the link is missing between protection of mental integrity and often mentioned concepts such as self-determination and autonomous decision-making. Lastly, human oversight

[80] Article 19, "EU: Better Human Rights Protections Needed in HLEG Guidelines on AI.", 2019. <https://www.article19.org/resources/eu-better-human-rights-protections-needed-in-hleg-guidelines-on-ai/>.
[81] Center for Democracy and Technology, "CDT's Comments to European Commission on Artificial Intelligence (AI HLEG)'s Draft Ethics Guidelines for Trustworthy AI.", 2020/ <https://cdt.org/insights/cdts-comments-to-european-commission-on-artificial-intelligence-ai-hlegs-draft-ethics-guidelines-for-trustworthy-ai>.

and accountability, which will be discussed below, are intimately linked concepts.[82] Nevertheless, the text does not make a direct and clear link between them, nor does it mention redress mechanisms for 'harms' caused by not complying with this requirement.

In order to comply with the *human agency and oversight* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (see Figure 3 below).

Respect for fundamental rights of individuals: How are you dealing with the effect of the application on the rights to safety, health, non-discrimination, and freedom of association?

a. We've performed a clear analysis in response to these principles and can provide details. (2)
b. We have partially/informally considered these principles but no specific details can be provided. (1)
c. We have not considered these issues yet. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Education and tutorials: Do you ensure that users are informed and capable of using the system correctly?
a. We provide complete in-system help (2),
b. We provide support through external materials, e.g. website. (1)
c. We do not provide user support. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Ease to deactivate/remove: How easy is it to deactivate or remove the system and data once users are no longer interested or need the system?
a. Very easy, either through clear instructions or automatically by the sunset clause. (2)
b. Instructions on how to deactivate or remove the system and data are unclear. (1)
c. There are no instructions or automated procedures to remove the system and the data. (0)
d. We consider that these issues are not applicable to our case (N/A)

Ease to access services without using the AI system: In the case of AI systems aimed to replace or complement public services, are there full non-system alternatives?
a. Yes, there is an easily accessible full non-system alternative. (2)
b. There is a partial alternative or access to the full alternative is unclear. (1)
c. There is no alternative to the AI system for this service. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 3: The abbreviated assessment list: human agency and oversight requirement[83]

[82] Renda A., "Europe." The Oxford Handbook of Ethics of AI, 1st ed., Oxford University Press, 2020.
[83] Dignum and others (n 41).

Technical robustness and safety

This requirement deals with four main issues: 1) security; 2) safety; 3) accuracy; and 4) reliability, fall-back plans and reproducibility. Technical robustness is closely linked to the principle of prevention of harm which requires that AI systems are developed with a preventative approach to risks. They should behave reliably while minimising unintentional and unexpected harm and preventing unacceptable harm. For AI systems to be considered secure, possible unintended applications of the AI system (e.g. dual-use applications) and potential abuse of the system by malicious actors such as data-targeted attacks (data poisoning[84]), model-targeted attacks (model flaws[85]), or software and hardware attacks should be taken into account. Steps should be taken to prevent and mitigate these risks which can include robust learning (redesigning the learning procedure of the AI systems or the algorithm or conducting explicit training against adversarial examples).[86] A media related illustration of a non-robust AI systems is the famous Chatbot Tay, which turned out after 16 hours of service to become racist. The system was learning from its interaction with other Twitter users and learned from their inappropriate comments which led to untended application of the human engagement project of Microsoft.[87]

Regarding safety, AI systems should also have safeguards that enable a fall-back plan. It must be ensured that the system will do what it is supposed to do without harming humans or the environment. Fall-back plans can be quite diverse: technical switching procedures or asking for a human operator before proceeding.[88]

Accuracy pertains to an AI system's ability to make correct judgements, predictions, recommendations, or decisions based on data or models. It is important that the system can indicate how likely these errors are and to design and think beforehand what the harm could be in case of materialisation of inaccurate predictions. To improve the accuracy, extensive testing on various data sets can help to identify edge cases that can arise.[89]

Reliability requires to scrutinising an AI system and to prevent unintended harms. Data sanitization is an approach recommended to increase the reliability of machine learning models.

[84] Defined as "deliberately introducing false data at the training stage of the mode", [in]: European Commission. Joint Research Centre., *Robustness and Explainability of Artificial Intelligence: From Technical to Policy Solutions.* (Publications Office 2020) <https://data.europa.eu/doi/10.2760/57493> accessed 22 August 2021.
[85] "It consists in taking advantage of the inherent weaknesses of the mathematical procedures involved in the learning process of the model", ibid.
[86] ibid., p.18.
[87] Vincent J, The Verge, 'Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day' (2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist.
[88] Koshiama A. and Engin Z., 'Algorithmic Impact Assessment: Fairness, Robustness and Explainability in Automated Decision-Making', (2019), Data for Policy 2019: Digital Trust and Personal Data (Data for Policy 2019) (DFP), London, https://doi.org/10.5281/zenodo.3361708.
[89] European Commission. Joint Research Centre. (n 84).

It asks to "clean the training data of all potential malicious content before training the model".[90] Reproducibility describes whether an AI experiment exhibits the same behaviour when repeated under the same conditions.

In order to comply with the *technical robustness and safety* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 4).

Security: Do you have user authentication in place to prevent risks such as access, modification, or disclosure of the data? Do you use unique and pseudo-random identifiers, renewed regularly and cryptographically strong?
a. Strong security elements are in place, e.g. user authentication, unique identifiers regularly renewed. We can provide further information. (2)
b. Some security features are in place. (1)
c. No security features are in place. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Ease to deactivate/remove: How easy is it to deactivate or remove the system and data once users are no longer interested or need the system?
a. Very easy, either through clear instructions or automatically by the sunset clause. (2)
b. Instructions on how to deactivate or remove the system and data are unclear. (1)
c. There are no instructions or automated procedures to remove the system and the data. (0)
d. We consider that these issues are not applicable to our case (N/A)

Open-source code: Is the development participatory and multidisciplinary? What kind of access to the code and development is there?
a. The code and development are open-source. (2)
b. The code is open-source code without the possibility of contributing. (1)
c. Non-open-source code. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 4: The abbreviated assessment list: technical robustness and safety[91]

Privacy and data governance

Privacy is enshrined in international human rights law[92] and strengthened by national data protection laws and jurisprudence. In Europe, privacy is considered as a fundamental right.[93]

[90] ibid.
[91] Dignum and others (n 41).
[92] Privacy is a fundamental human right recognized in the UN Declaration of Human Rights, the International Convenant on Civil and Political Rights and in many other international and regional treaties.
[93] The right to privacy or private life is enshrined in the European Convention of Human Rights (Article 8) and the European Charter of Fundamental Rights (Article 7).

How (the right to) privacy can be affected by AI systems has been a prominent theme in many documents. Notably, the EPRS report on "The ethics of artificial intelligence: Issues and initiatives" notes that "AI will have profound impacts on privacy in the next decade."[94] It lists numerous ways in which AI systems interplay with privacy and data rights. One example being the use of machine learning which can extract information from data and discover new patterns and is able to turn seemingly innocuous data into sensitive, personal data.[95] Moreover, personal data in the training set can, in certain cases, be reconstructed from a model. Mitrou identified the following tendencies which may have adverse implications for data processing and protection when combined with AI.[96] First, the paradigm of collection of "all data" or "as much data as possible" to be able to further learn and analyse. Second, re-purposing or multi-purposing of data that is generated in a specific context or activity but may be used and analysed for another, initially unknown purpose. Undoubtedly, whether for "training purposes" or as part of their deployment, AI involves the processing of personal information, which is subject to the applicable legal and ethical framework.

The AI HLEG Guidelines start with the realisation that prevention of harm to privacy necessitates adequate data governance that covers the quality and integrity of the data used, its relevance in light of the context in which the AI systems is deployed, its access protocols and the capability to process data in a manner that protects privacy. Then, the Guidelines point out that privacy and data protection require that AI systems must guarantee privacy and data protection throughout a system's entire lifecycle. This includes the information initially provided by the user, as well as the information generated about the user over the course of their interaction with the system (e.g. outputs that the AI system generated for specific users or how users responded to particular recommendations). Moreover, to allow individuals to trust the data gathering process, it must be ensured that data collected about them will not be used to unlawfully or unfairly discriminate against them. In the same vein, Mitrou points out that respecting privacy and data protection laws is not simply a matter of (demonstrating) compliance with the legal framework. The acceptance and consequently, the use of AI is highly dependent on the trust of the users. The confidence that a user's informational privacy – the capacity of an individual to control information about herself - is protected, is one of the prerequisites of this trust.[97]

---

[94] European Parliament. Directorate General for Parliamentary Research Services., *The Ethics of Artificial Intelligence: Issues and Initiatives.* (Publications Office 2020) <https://data.europa.eu/doi/10.2861/6644> accessed 21 April 2021.

[95] ibid.

[96] Mitrou L., 'Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) "Artificial Intelligence-Proof"?' [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3386914> accessed 21 April 2021.

[97] ibid.

Another important caveat of the privacy and data protection requirements is the quality and integrity of data. More specifically, the quality of the data sets used is paramount to the performance of AI systems. When data is gathered, it may contain socially constructed biases, inaccuracies, errors and mistakes. The problem with quality and representation of the training data, especially those in publicly available datasets and databases, is well recognized in the academic literature. As mentioned by Raji and others, "privacy and consent violations in the dataset curation process often disproportionately affect members of marginalized communities. Benchmark dataset curation frequently involves supplementing or highlighting data from a specific population that is underrepresented in previous dataset."[98] To illustrate, the authors point out that "CelebSET sourced from IMDB-WIKI contained a significant demographic bias that can be seen in the distribution of meta-data labels."[99] There are a number of studies showing that in the publicly available datasets certain groups are highly underrepresented. The problem is even more visible when it comes to the intersectional identities.[100] To this end, it is likely that using such data could lead to algorithmic results being biased and discriminatory. The HLEG Guidelines note that "this needs to be addressed prior to training with any given data set" (own emphasis).[101]

In addition, the Guidelines point out that the integrity of data must be ensured. Feeding malicious data into an AI system may change its behaviour, particularly with self-learning systems. To this end, processes and data sets used must be tested and documented at each step such as planning, training, testing and deployment. Importantly, this also applies to AI systems that were not developed in-house but acquired elsewhere.

Finally, the Guidelines require that in any given organisation that handles individuals' data, data protocols governing data access should be put in place. These protocols should outline who can access data and under which circumstances. Only duly qualified personnel with the competence and need to access an individual's data should be allowed to do so.

It is important to note that the elements of the privacy and data governance requirement are already substantiated by binding EU data protection legislation, in particular the GDPR.

The GDPR contains important rights for users relating to any processing of their personal data as well as obligations for data controllers and processors. The applicability of the GDPR results in the obligation of data controllers (and processors) to comply with its requirements that relate

[98] Raji I.D. and others, 'Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing' [2020] arXiv:2001.00964 [cs] <http://arxiv.org/abs/2001.00964> accessed 27 July 2021.
[99] ibid.
[100] Buolamwini J., Gebru .T., 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' 15. More information about the Gender Shades project which evaluated the accuracy of AI powered gender classification products can be found here: http://gendershades.org/overview.html.
[101] 'High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy AI" (European Commission 2019)' (n 38).

to the legal ground of processing, the data protection principles, the respect for the rights of the data subjects and the obligations in relation to organizing, ensuring and demonstrating compliance with the GDPR (accountability, DPIA). "Privacy by design," also known as data protection by design, is an obligation stemming from Article 25 of the GDPR to integrate considerations of data privacy into the construction of an AI system and the overall lifecycle of the data. According to the GDPR data controllers must "implement appropriate technical and organisational measures..." during the design and implementation stage of data processing "to protect the rights of data subjects."[102]

The detailed analysis of the applicability of the GDPR to AI systems falls outside the scope of this deliverable and will be subject of a detailed analysis in Deliverable D4.3 "Initial analysis of the legal and ethical framework of trusted AI".

To see how the requirement of privacy and data governance is further operationalized in EU (proposed) legislation, in particular in Data Governance Act and AI Regulation, see Section 4.3 and 4.1.3 respectively.

In order to comply with the *privacy and data governance* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 5).

[102] Art. 25 of the GDPR.

Privacy and data protection: Is data collection compliant with the General Data Protection Regulation (GDPR) and does it respect the privacy of the individual? Note that A Data Protection Impact Assessment (DPIA) must be carried out before the deployment of any system.
a. The purpose of the AI system and the mechanisms to assess its usage are clearly defined and compliant with GDPR, a DPIA has been performed and privacy of individuals is guaranteed. We can provide further information. (2)
b. We have only done a partial/informal analysis and/or not all aspects of data and privacy protection are clear. (1)
c. We cannot guarantee privacy and data protection. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Data management: Do you comply with the data-minimization principle, i.e. usage of local and temporary storage and encryption, based on principles of data protection by design? Do you ensure that only strictly necessary data are captured and processed?
a. We use local and temporary storage and data encryption methods. We only collect and process strictly necessary data. We can provide further information. (2)
b. We partially comply with the above, and some documentation can be provided. (1)
c. We do not comply with these data management aspects. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Ownership: Is the ownership of the resource clear?
a. Ownership of the resource (including code, data, use) is clear and explicit. (2)
b. Some ownership aspects are made clear. (1)
c. Ownership information for the resource and related code or data is unavailable. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 5: The abbreviated assessment list: privacy and data governance requirement[103]

Transparency

This requirement is closely linked with the principle of explicability (see 2.2.1.2 above) and encompasses transparency of elements relevant to an AI system: the data, the system and the business models.

The AI HLEG Guidelines note that the data sets and the processes that yield the AI system's decision, including those of data gathering and data labelling as well as the algorithms used, should be documented to the best possible standard to allow for traceability. This also applies to the decisions made by the AI system. Such traceability would enable the identification of the reasons why an AI-decision was erroneous which, in turn, could help prevent future mistakes. In short, traceability facilitates both auditability and explainability.

[103] Dignum and others (n 41).

The second caveat of the transparency requirement is the explainability. Chatila and others note that "explainability is at the heart of Trustworthy AI and must be guaranteed for developing AI systems aimed at empowering and engaging people, across multiple scientific disciplines and industry sectors."[104] The exact meaning of "explainability" and its practical operationalization has been subject of many academic debates. From the machine learning point of view, different interpretable/explainable models are developed to understand the mathematical processes behind decisions.[105] From a legal and ethical point of view, however, explainability can be defined as meaningful insights on how a particular decision is made.[106] As argued by Bibal and others, it is not necessarily required to provide an interpretable representation of a mathematical model.[107] Most important is an explanation that can make the decision meaningful for an individual, i.e. so that the decision makes sense to them. It follows from the AI HLEG Guidelines that explainability should be adapted to the level of expertise and understanding of the individual. In AI HLEG's own words, "such explanation should be timely and adapted to the expertise of the stakeholder concerned (e.g. layperson, regulator or researcher)."[108]

Moreover, there are different meanings of explainability; some may relate to the overall process of decision making, other to the final decision and their scope depends on the impacts that the algorithmic decision has on users' life. In AI HLEG's view, explainability concerns the ability to explain both the technical processes of an AI system and the related human decisions (e.g. application areas of a system). Technical explainability requires that the decisions made by an AI system can be understood and traced by human beings. Moreover, the Guidelines note that whenever an AI system has a significant impact on people's lives, it should be possible to demand a suitable explanation of the AI system's decision-making process. In addition, explanations of the degree to which an AI system influences and shapes the organisational decision-making process, design choices of the system, and the rationale for deploying it, should be available (hence ensuring business model transparency).

It is important to note that European and national laws already contain several biding legal obligations on explainability.[109] Some rules apply generally, to all types of decision-making, while other rules, often stricter, apply specifically to "automated decision-making". Wachter and

[104] Chatila R. and others, 'Trustworthy AI' in Bertrand Braunschweig and Malik Ghallab (eds), *Reflections on Artificial Intelligence for Humanity*, vol 12600 (Springer International Publishing 2021) <http://link.springer.com/10.1007/978-3-030-69128-8_2> accessed 26 July 2021.
[105] Bibal A., Frénay B., 'Interpretability of Machine Learning Models and Representations: An Introduction' 7.
[106] Bibal A. and others, 'Legal Requirements on Explainability in Machine Learning' (2021) 29 Artificial Intelligence and Law 149.
[107] ibid.
[108] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019. (n 31).
[109] Bibal and others (n 106).

others propose the following classification of what one may mean by an 'explanation' of automated decision-making.[110] Two kinds of explanations are possible, depending on whether one refers to: system functionality, i.e. the logic, significance, envisaged consequences, and general functionality of an automated decision-making system, e.g. the system's requirements specification, decision trees, pre-defined models, criteria, and classification structures; or to specific decisions, i.e. the rationale, reasons, and individual circumstances of a specific automated decision, e.g. the weighting of features, machine-defined case-specific decision rules, information about reference or profile groups. Furthermore, one can also distinguish between an ex-ante explanation (i.e. prior to an automated decision-making taking place) and an ex-post explanation (i.e. after an automated decision has taken place). [111] Focus of many legal scholars has been on the meaning of explainability from the data protection law point of view. The core debate has primarily focused on whether or not the GDPR creates a *right* to explanation of an algorithmic decisions.[112] This debate falls outside the scope of this Deliverable. Importantly, the focus of many legal scholars has been on how the legal requirements on explainability could be interpreted and applied in machine learning. [113] Hamon and others used a COVID-19 use case scenario to assess the feasibility of legal requirements on algorithmic explanations.[114] They concluded that the use of complex deep learning models in AI applications, such as in COVID-19 detection, makes it hard to reconcile with the existing EU data protection law requirements, especially with regards to human legibility of explanations for non-expert data subjects.

More recently, the focus on explainability has been criticized. Edwards and Veale argue that it is possible that in some cases transparency or explanation rights may be overrated or even irrelevant – the problem which is often referred to as "transparency fallacy." [115] In fact, in many cases what the data subject wants is not an explanation—but rather for the disclosure, decision

[110] Wachter S., Mittelstadt B., and Floridi L., 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76.
[111] ibid.
[112] See: Goodman B. and Flaxman S., 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' [2016] arXiv:1606.08813 [cs, stat] <http://arxiv.org/abs/1606.08813>; Wachter S., Mittelstadt B. and Floridi L., 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76; Edwards L. and Veale M., 'Slave to the Algorithm? Why a "right to an Explanation" Is Probably Not the Remedy You Are Looking For' (LawArXiv 2017) preprint <https://osf.io/97upg> accessed 21 April 2021; Selbst A. D. and Powles J., 'Meaningful Information and the Right to Explanation' (2017) 7 International Data Privacy Law 233.
[113] Bibal and others (n 106).
[114] Hamon R. and others, 'Impossible Explanations?: Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445917> accessed 24 May 2021.
[115] Edwards and Veale (n 112).

or action simply not to have occurred.[116] Along the same vein, Hildebrandt argues that the discussion on explainability must not stand in the way of the question whether a decision is legally *justified*.[117] For instance, the conviction of a defendant based on the predictive accuracy of an algorithm, even if "explained", is only possible if there are the legal grounds that justify such conviction.[118] Such justification does not concern the explanation of how an AI system works, but reasons as provided by law.

This is not to say that the requirement of explainability should be set aside. On the contrary, there is an ever-growing need for inter-disciplinary research on how the legal and ethical explainability requirements can be applied in ML practice. In addition, more research is needed to see how to accommodate explainability alongside other ethical requirements for trustworthy AI. In particular, as noted by the AI HLEG Guidelines, trade-offs might have to be made between enhancing a system's explainability (which may reduce its accuracy) and increasing its accuracy (at the cost of explainability).

The last dimension of the transparency requirement is the fact that humans should be informed that they are interacting with an AI system. They should also have the option to have a human interaction instead. Beyond this, the AI HLEG Guidelines note that the AI system's capabilities and limitations should be communicated to AI practitioners or end-users in a manner appropriate to the use case at hand. This could encompass communication of the AI system's level of accuracy, as well as its limitations.

In order to comply with the *transparency* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 6).

[116] ibid.
[117] Hildebrandt M., 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20 Theoretical Inquiries in Law 83.
[118] ibid.

Transparency rights: Do you include the right of users to:

(i) be notified that their data is being processed/collected,
(ii) access information on which personal data are collected,
(iii) control their own data,
(iv) access explanations of results produced by the system,
(v) be informed of who, when and how the system can be audited.

a. All of the above are fulfilled. (2)
b. Only some of the above are fulfilled or partially addressed. (1)
c. We cannot guarantee any transparency aspects. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Openness over Data governance: How open is Data governance?
a. Open data governance. (2)
b. Intermediate openness of data governance. (1)
c. Private/opaque settings. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Legislation and Policy: Are there explicit legislation and/or other policies relevant to your system/resource?
a. The system is covered by an explicit clear, legal framework or sectorial formal policies, and we address these explicitly. (2)
b. We are aware of policy partially relevant to our system and address these sufficiently. (1)
c. We are not aware of any relevant legislation or policy and do not address these. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Design Impact Assessment and Open Development Process:
How explicit is the design process leading to this resource?
a. Explicit information on the design process is available, including a clear description of aims and motivation, stakeholders, public consultation process and impact assessment. (2)
b. Some information on the design process, aims and motivation, and impact assessment is available. (1)
c. There is no information on the design and impact of the resource. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 6: The abbreviated assessment list: transparency requirement[119]

Diversity, Non-discrimination and Fairness

The fifth ethical requirement's inclusion and diversity components are closely linked with the principle of fairness, as the principle prescribes that the development, deployment and use of AI systems must be fair by ensuring equal and just distribution, safeguarding from unfair bias, providing equal access opportunities, respecting the principle of proportionality between means

[119] Dignum and others (n 41).

and ends, and balancing competing interests and objectives. This requirement, therefore, could be fulfilled by enabling inclusion and diversity through the entire life cycle of the AI system in the following ways: (i) consideration and involvement of all affected stakeholders in the entire process; (ii) guarantee equal access through inclusive design; and (iii) ensuring equal treatment.

The Guidelines interpret this ethical requirement in three sub-sections: (1) Avoidance of unfair bias, (2) Accessibility and Universal Design, and (3) Stakeholder Participation.

Avoidance of Unfair Bias

Datasets used by AI systems, whether in training or operation, may suffer from inclusion of inadvertent historical bias, incompleteness, and bad governance. Hence, continuation of such bias may lead to unintended (in)direct prejudice and discrimination against certain groups or people; and potentially exacerbating prejudice and marginalization. As scholars and civil society draw attention, it is inevitable that AI developers may have their own values and unconscious bias planted in their psyche by society's outdated practices. That is why it is crucial that they are aware of algorithms with historically unfair prejudices during the development phase of the AI systems.[120] AI organisations should also invest in ways to identify, address, and mitigate unfair biases at every stage of the development process, while ensuring that accurate and representative sample data is collected, analysed, and used.[121] Additionally, data that is being used should be representative of the target population and should be as inclusive as possible, by not only focusing on exclusion issues but also promoting active inclusion, diversity hiring, and usage of fairness-aware data mining algorithms[122] in the development and design of AI. [123] Lastly, organisations using AI need to ensure that the outcomes of AI decisions are reversible, especially when there is a harm caused by those.

Thus, some principles that are not mentioned explicitly or in detail in the Guidelines should also be taken into consideration to interpret the sub-requirement much more comprehensively. These are the principles of consistency, equality, inclusion, equity, plurality, and reversibility.

To sum up, the mechanisms below, along with aforementioned tools, are recommended by the Guidelines to counteract the effects of unfair bias:

1. Removing identifiable and discriminatory bias in the collection phase, where possible;

[120] Latonero M, 'Governing Artificial Intelligence: upholding human rights & dignity', Data&Society, https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf
[121] ibid.
[122] FATML, "Principles for accountable algorithms and a social impact statement for algorithms", 2016. https://www.fatml.org/resources/principles-for-accountable-algorithms.
[123] Gilburt, B., "Women leading in AI: 10 principles of responsible AI", Towards Data Science, 2019. https://towardsdatascience.com/women-leading-in-ai-10-principles-for-responsibleai-8a167fc09b7d.

2. Putting in place oversight mechanisms that would: (1) counteract with AI systems suffering from unfair bias, and (2) analyse and address the system's purpose, constraints, requirements, and decisions in a clear and transparent manner.
3. Hiring and encouraging hiring from diverse backgrounds, cultures, and disciplines to ensure diversity of opinions.

Accessibility and Universal Design

According to the Guidelines, AI systems, particularly in business-to-consumer domains, should be user-centric and designed in a way that allows all people to use AI products or services, regardless of their age, gender, abilities, or characteristics. Thus, it should be ensured that AI systems should not have a one-size-fits all approach, and they should consider Universal Design principles. These principles will allow AI systems to follow relevant accessibility standards, while ensuring that technologies developed are fair and accessible among a diversity of user groups within society,[124] especially among those that currently lack such access.[125] Moreover, ensuring accessibility also presents a crucial importance for persons with disabilities present in all societal groups. Additionally, individuals should be able to access the explanations when decisions are made about them, and these explanations should be easily accessed, free of charge, and user friendly.[126] Thus, wherever possible, AI systems should adapt open data to ensure access and transparency.[127] Conclusively, AI systems respecting this sub-requirement would enable equitable access and active participation of all people in existing and emerging computer-mediated human activities and with regard to assistive technologies.

Stakeholder Participation

The Guidelines do not provide extensive information on this sub-requirement. Nevertheless, civil society and scholars urge AI developers to consider the range of social and cultural viewpoints, while attempting to prevent societal homogenization of behaviour and practices.[128] There should be consistent, repeated, and regular discussions with end users and stakeholders that maybe affected. This includes creating a multi-stakeholder dialogue and incorporating the viewpoints from a wide-range of people such as women, underrepresented groups, and marginalised communities at every stage of AI applications.[129] Moreover, AI developers should

[124] Smart Dubai, "Artificial intelligence principles and ethics", 2019. https://www.smartdubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf.
[125] AI Now Institute, "The AI now report: the social and economic implications of artificial intelligence technologies in the near-Term", 2016. https://ainowinstitute.org/AI_Now_2016_Report.pdf.
[126] ibid.
[127] NSTC, "The national artificial intelligence research and development strategic plan", 2016. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf.
[128] University of Montreal, "Montreal declaration for a responsible development of artificial intelligence", 2017. https://www.montrealdeclaration-responsibleai.com/.
[129] Leaders of the G7 , "Common vision for the future of artificial intelligence", 2018. https://www.mofa.go.jp/files/000373837.pdf.

solicit regular feedback even after deployment and set up mechanisms for stakeholder participation in the long run; for instance, by ensuring that workers are informed and consulted, and being able to participate throughout the whole process of implementing AI systems at organizations. To conclude, AI should not lead discrimination against people or groups of people based on gender, race, culture, religion, age, ethnicity, etc.[130] There should be opportunities for challenge, redress, remedy, and reversibility regarding outcomes resulting from usage of AI systems. This of course starts with identifying sexist, misogynistic, gender-biased, racist, and similar harms resulting from discriminatory practices.[131] For that reason, the next section discusses gender inequality and other algorithmic discrimination practices that have been recognized in the EU, by providing a detailed information on the special report issued by the EC.

In order to comply with the *diversity, non-discrimination, and fairness* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 7).

Inform on how it is respecting fundamental rights of individuals: How are you dealing with the effect of the application on the rights to safety, health, non-discrimination, and freedom of association?
a. We've performed a clear analysis in response to these principles and can provide details. (2)
b. We have partially/informally considered these principles but no specific details can be provided. (1)
c. We have not considered these issues yet. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Accessibility: Can your app/system/resource be used by all regardless of demographics, language, disability, digital literacy, and financial accessibility?
a. This resource is fully accessible, and we can provide information on accessibility accommodations. (2)
b. This resource partially complies with accessibility requirements. (1)
c. This resource is not accessible to all. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 7: The abbreviated assessment list: diversity, non-discrimination, and fairness[132]

Report on Algorithmic Discrimination in Europe

It is also worth to mention that the specific risks posed by AI algorithms in terms of gender inequality have been recognized by the Commission's recent Gender Equality Strategy 2020-

[130] Cerna Collectif, "Research ethics in machine learning", 2018. https://hal.archives-ouvertes.fr/hal-01724307/document.
[131] World Wide Web Foundation, "Artificial intelligence: Open questions about gender inclusion", 2018. http://webfoundation.org/docs/2018/06/AI-Gender.pdf.
[132] Dignum and others (n 41).

2025.[133] With the aim of addressing gender discrimination, the "Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non Discrimination Law" report examines the current gender equality and non-discrimination legislative framework in place in the EU in light of algorithmic discrimination.[134] In the context of EU gender equality and non-discrimination law, algorithmic discrimination refers to discrimination based on one of the six grounds explicitly listed in and protected under Article 19 TFEU, that is sex, race or ethnic origin, disability, sexual orientation, religion or belief and age. The report details the various phases in which discrimination can creep into algorithms. From design to use, and from planning to development and decision-making, bias can impact algorithms in several ways. The specific problems that arise in relation to algorithmically supported decisions are classified in the following types:

- The biases in the data (revealing the stereotypes and cognitive biases pervading the humans' visions and decisions, describing historically consolidated patterns of discrimination structuring society);
- The discriminatory effects of algorithms (how algorithms might reify and further enact discriminatory correlations, implicitly and wrongly considered causations when such an algorithm drives the decisions);
- Transparency and lack of information (how to explain how an algorithm made such a decision, do judges or citizens are given formal access to the inner-workings of the algorithm);
- Responsibility and accountability (many different parties are involved in the design, commercialization and use of algorithms);
- The gender digital gap in Europe (stark overrepresentation of males not pertaining to minorities in STEM education and professions).

After classifying the problems, the report then analyzes how the current non-discrimination EU legislative framework can adequately capture and redress algorithmic discrimination. For example, algorithmic profiling based on granular analysis of personal and behavioural data indeed entails heightened risks of intersectional discrimination, a type of discrimination that the Court of Justice has so far failed to adequately recognise. Since the media is excluded from Directive 2004/113/EC, these types of representational intersectional and sex discrimination are out of reach of EU gender equality law. A similar scope issue arises in relation to sex discrimination in online advertising, which is excluded from the scope of the directive. Harmful stereotyping and prejudices could pervade algorithms used to determine the distribution of ads and ultimately access to goods and services (education, housing, etc.). In particular, the

[133] European Commission, "Gender Equality Strategy'', 2020. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en.

[134] Directorate-General for Justice and Consumers (European Commission), and others Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non Discrimination Law. Publications Office of the European Union, 2021.

exclusion of education from the scope of Directive 2004/113/EC is problematic in light of the under-representation of women in STEM fields and curricula related to IT and software development. This lack of voice of women and minority groups in algorithmic design has clear repercussions in terms of biased algorithmic design leading to discrimination. The report underlines the lack of binding positive action measures on the side of public authorities in Europe. Indeed, policy-making bodies appear to be wary of such positive action, in particular quota policies, although such temporary measures could dramatically contribute to closing the gender digital gap in the future. Finally, the report proposes its own integrated set of legal, knowledge-based and technological solutions to the problem of algorithmic discrimination. For instance, it highlights for the above issue on the lack of diversity in IT that in addition to favoring a diverse IT workforce, AI professionals and data scientists need to be specifically trained to recognise, avoid, and test for these biases when designing algorithmic applications. One way to conduct such training would be by adapting university curricula and vocational and professional training to include digital humanities, social sciences and ethics components.

Societal and environmental well-being

The requirement of societal and environmental well-being is in line with the UN Sustainable Development Goals and pushes forward AI systems towards benefits to all human beings including future generations. Concerning the environmental well-being, the Guidelines encourage sustainability and ecological responsibility of AI systems. Through their whole life cycle, AI systems should be designed in the most environmentally friendly way possible. An entire range of aspects would need to be considered for instance: the type of energy used (renewal, fossil fuels), the cloud computing infrastructure, the energy consumption of the AI system, the components used, the recycling aspects of the AI system's components waste. Social impact of the AI systems should also be carefully assessed including the individuals' physical and mental well-being as their exposure to AI systems covers all aspects of their lives and can mislead individuals. When it comes to the societal impact, the Guidelines request that the effects of AI systems on political institutions, democracy and society are carefully monitored. Attention should be drawn to these AI systems which could hinder the democratic processes. For this aspect, there is a need to look at the horizon and assess the AI system's impact from a more global perspective.

RESEARCH ON AI AND ENVIRONMENT

There are two sides of the research on environment and AI. It is a domain where balance between AI and the environment is delicate to reach. Indeed, AI can be used to help protecting

the environment, but its extensive use can damage and harm it.[135] The research focuses on the one hand, on how AI can help mitigate or counter the impacts of climate change such as decoupling economic growth from rising carbon emissions and environmental degradation.[136] This goes from better energy management, traffic prediction, freight allocation, precision agriculture enabling better use of resources like land and water, better weather forecasting, and optimal heating and cooling of buildings, detection of illegal environmental activities, better management of environmental disasters.[137] On the other hand, research is done on how to reduce the negative environmental impact of AI systems themselves, during training and deployment. The metrics to assess the environmental impact of AI are being debated and there is a need to define (environmental) well-being.[138] Researchers also acknowledge the challenges emerging from AI research as "energy-efficient AI may be less prestigious because it may not attain the same levels of accuracy and performance as AI that is unrestricted in how much energy it uses."[139] Research is progressing on the energy consumption necessary for training and developing AI systems. For instance, a study from the MIT showed how training a single AI model can emit as much carbon as five cars in their lifetimes.[140]

Researchers at the Montreal AI Ethics Institute, McGill University, Carnegie Mellon, and Microsoft have developed a framework designed to quantify the environmental and social impact of AI.[141] The framework will permit to cut contributions to the carbon footprint while at the same time addressing trustworthiness and data sovereignty. A semi-automated certification process is also part of the framework and would permit users to assess the state of an AI system in comparison with others. Deploying such framework at scale will enable consumers, academics, and investors to demand more transparency on the social and environmental

[135] Vinuesa, R., Azizpour, H., Leite, I. and others The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020). https://doi.org/10.1038/s41467-019-14108-y

[136] Branch, 'AI and Climate Change: The Promise, the Perils and Pillars for Action', https://branch.climateaction.tech/issues/issue-1/ai-and-climate-change-the-promise-the-perils-and-pillars-for-action/.

[137] Montreal AI Ethics Institute, 'State of AI Ethics Reports' (July 2021),p.117, https://montrealethics.ai/volume5/; McKInsey Global Institute, 'Notes from the AI Frontier applying AI for social good, Discussion Paper' (2018) mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf

[138] Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the Carbon Emissions of Machine Learning. ArXiv, abs/1910.09700.; Environmental Intelligence: Applications of AI to Climate Change, Sustainability, and Environmental Health (stanford.edu).

[139] Gupta A., Lanteigne C. and Kingsley S., (2020), 'SECure: A Social and Environmental Certificate for AI Systems', 'Computers and Society', https://arxiv.org/ftp/arxiv/papers/2006/2006.06217.pdf, p. 4.

[140] MIT Technology Review, 'Training a single AI model can emit as much carbon as five cars in their lifetimes' (2019), https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

[141] Gupta A., Lanteigne C. and Kingsley S., (2020), 'SECure: A Social and Environmental Certificate for AI Systems', 'Computers and Society', https://arxiv.org/ftp/arxiv/papers/2006/2006.06217.pdf.

impacts of AI and lead to better informed choices to steer progress towards AI systems showing less negative impact to environmental and societal wellbeing.[142]

Responsible AI Licenses (RAIL) is another way to steer progress towards positive impact for environment and society as it "empowers developers to restrict the use of their AI technology in order to prevent irresponsible and harmful applications".[143] These licenses include "clauses for restrictions on the use, reproduction, and distribution of the code for potentially harmful domain applications of the technology".[144]

Research on AI and societal well-being

Research conducted in the field of AI and societal well-being shows how AI can enable the achievement of societal well-being goals but can also be used to inhibit these targets. This is why having strong regulatory insight and oversight is needed to ensure the AI development is taking the good path.[145] There is a need for a definition on societal well-being sometimes also called "community well-being", development of indicators for assessing the impact of AI on societal well-being, research on interdependency between the process of developing and deploying AI and impacts on community well-being. [146] The Mc Kinsey Global Institute compiled a collection of 160 AI social-impact use cases which can contribute to improve all of the SDG and help society.[147] The research team identified bottlenecks which could limit AI uptake and benefit to society. Even if AI brings a lot of promises its benefits for societal wellbeing could be circumvented by misuses of the public authorities or private stakeholder deploying them or by unintentional harm caused by the AI system. For this reason, effective mitigation strategies need to be put in place to ensure the positive societal impact of the AI system. Data are the crucial enabler for the uptake of AI systems for the societal wellbeing. Indeed, data from users, from public or private entities are needed to fuel its positive application; therefore, research improving data accessibility, quality and diversity for social-impact cases is essential.[148] Researchers have observed a shortage of experienced AI professionals in the social sector. Increasing investment in education programmes and PhD grants for societal AI will be a key enabler for progress in this field.[149] Researchers also observed that one of the drawbacks of AI

[142] Wiggers K, Venture Beat, Researchers propose framework to measure AI's social and environmental impact (2020), https://venturebeat.com/2020/06/12/researchers-propose-framework-to-measure-ais-social-and-environmental-impact.
[143] Responsible AI Licenses (RAIL), https://www.licenses.ai/
[144] ibid.
[145] Vinuesa R. and others, 'The Role of Artificial Intelligence in Achieving the Sustainable Development Goals' (2020) 11 Nature Communications 233.
[146] Musikanski, L., Rakova, B., Bradbury, J. *and others* Artificial Intelligence and Community Well-being: A Proposal for an Emerging Area of Research. *Int. Journal of Com. WB* 3, 39–55 (2020). https://doi.org/10.1007/s42413-019-00054-6
[147] McKInsey Global Institute, op.cit.
[148] ibid.
[149] ibid.

development for societal well-being is that progress in this field is often governed and based on the value and needs of the nations in which the AI system is being developed. This means that in States without a good democratic control, AI could enable nationalism, hate towards minorities, and bias election outcomes.[150] Additionally, without a framework considering the societal impact of AI systems, it was observed that AI even if developed in democratic countries was very much driven by capitalist aims that do not align with societal values. Better collaboration between AI researchers and application-domain experts is also part of the solution to establish interdisciplinary partnerships for socially good AI. [151] When it comes to AI for social good, a publication maps the actors active in the field, existing studies, and results.[152] It is however underlined that it is not easy to achieve lasting impact in this field; early collaboration of experts on various aspects of AI is recommended, and guidelines on those are established.[153]

As a conclusion for this section, research is needed to understand what the benefits for AI and society are, how to achieve them, what the risks are, and how they can be mitigated. As provided by the "The role of artificial intelligence in achieving the Sustainable Development Goals" report, "regulatory oversight should be preceded by regulatory insight, where policymakers have sufficient understanding of AI challenges to be able to formulate sound policy. Developing such insight is even more urgent than oversight, as policy formulated without understanding is likely to be ineffective at best and counterproductive at worst."[154]

In order to comply with the *societal and environmental well-being* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 8).

[150] Vinuesa and others (n 147).
[151] Tomašev, N., Cornebise, J., Hutter, F. and others AI for social good: unlocking the opportunity for positive impact. Nat Commun 11, 2468 (2020). https://doi.org/10.1038/s41467-020-15871-z
[152] ibid.
[153] ibid.
[154] Vinuesa and others (n 147).

Education and tutorials: Do you ensure that users are informed and capable of using the system correctly?
a. We provide complete in-system help (2), or
b. We provide support through external materials, e.g. website. (1)
c. We do not provide user support. (0) d. We consider that these issues are not applicable to our case. (N/A)

Ease to access services without using the AI system: In the case of AI systems aimed to replace or complement public services, are there full non-system alternatives?
a. Yes, there is an easily accessible full non-system alternative. (2)
b. There is a partial alternative or access to the full alternative is unclear. (1)
c. There is no alternative to the AI system for this service. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Design Impact Assessment and Open Development Process: How explicit is the design process leading to this resource?
a. Explicit information on the design process is available, including a clear description of aims and motivation, stakeholders, public consultation process and impact assessment. (2)
b. Some information on the design process, aims and motivation, and impact assessment is available. (1)
c. There is no information on the design and impact of the resource. (0) d. We consider that these issues are not applicable to our case. (N/A)

Figure 8: The abbreviated assessment list: societal and environmental well-being requirement[155]

Accountability

According to AI HLEG's Guidelines, the requirement of accountability complements the above requirements, and is closely linked to the principle of fairness. It necessitates that mechanisms be put in place to ensure responsibility and accountability for AI systems and their outcomes, both before and after their development, deployment and use.[156]

The EP study "The ethics of artificial intelligence: Issues and initiatives" provides that "accountability ensures that if an AI makes a mistake or harms someone, there is someone that can be held responsible, whether that be the designer, the developer or the corporation selling the AI".[157] Importantly, as noted by Caplan and others, humans are the arbiters of the inputs, design of the system, and outcomes of an algorithm.[158] This is why, "critically, algorithms do not make mistakes, humans do".[159] Bryson notes that the extent to which transparency and

[155] Dignum and others (n 41).
[156] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).
[157] European Parliament. Directorate General for Parliamentary Research Services. (n 94).
[158] Caplan R. and others, 'Algorithmic Accountability: A Primer' (Data & Society 2018) <https://datasociety.net/output/algorithmic-accountability-a-primer/> accessed 12 February 2019.
[159] ibid.

accountability should be required is also a design decision which legislators, courts, and regulators have to make when designing a regulatory framework.[160]

The second component of the accountability requirements is auditability. Auditability entails the enablement of the assessment of algorithms, data and design processes. Importantly, AI HLEG notes that this does not necessarily imply that information about business models and intellectual property related to the AI system must always be openly available. However, especially in the context of applications affecting fundamental rights, including safety-critical applications, AI systems should be able to be independently audited and that evaluation reports by internal and external auditors shall be available.

The question raises how to ensure algorithmic auditability in practice. The practical implementation of algorithmic auditability has been a focus of some scholars. Brown and others define ethical algorithm audits "as assessments of the algorithm's negative impact on the rights and interests of stakeholders, with a corresponding identification of situations and/or features of the algorithm that give rise to these negative impacts."[161] They propose one way to operationalize high-level ethical analyses of algorithms by suggesting an auditing instrument which translates those ethical analyses into practical steps.[162] In the same vein, in their paper titled "Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing", Raji and others introduce a framework for algorithmic auditing that supports AI system development end-to-end, to be applied throughout the internal organization development lifecycle.[163] The proposed auditing framework is intended to contribute to closing the *accountability gap* in the development and deployment of large-scale AI systems by embedding a robust process to ensure audit integrity.[164] In the authors' view, an initial internal audit framework should encompass five distinct stages— Scoping, Mapping, Artifact Collection, Testing and Reflection (SMACTR). They all have their own set of documentation requirements.

The third component of the accountability requirement is the minimisation and reporting of negative impacts. The potential negative impacts of AI systems must be identified, assessed, documented and minimised. Moreover, due protection must be available for whistle-blowers, NGOs, trade unions or other entities when reporting legitimate concerns about an AI system. As noted by Caplan, currently, journalists are an important watchdog for algorithmic bias. They often use reverse-engineering to probe what's inside the black box or work collaboratively with

[160] Bryson J.J., 'The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation' 34.
[161] Brown S., Davidovic J. and Hasan A., 'The Algorithm Audit: Scoring the Algorithms That Score Us' (2021) 8 Big Data & Society 205395172098386.
[162] ibid.
[163] Raji I.D. and others, 'Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing' [2020] arXiv:2001.00973 [cs] <http://arxiv.org/abs/2001.00973> accessed 11 August 2021.
[164] ibid.

academics and whistle-blowers.[165] In AI HLEG's view, the use of impact assessments (e.g. red teaming or forms of Algorithmic Impact Assessment) both prior to and during the development, deployment and use of AI systems can be helpful to minimise negative impact. These assessments must be proportionate to the risk that the AI systems pose.

In particular, the concept of Algorithmic Impact Assessments (AIAs) has received a good deal of attention as possible tool to mitigate algorithmic harms and problems of algorithmic discrimination, bias, unfairness and accountability gap. Selbst notes that the AIA proposals can be put into three categories:

1) models based on National Environmental Policy Act (NEPA).[166] The NEPA model implies a highly detailed impact assessment, transparent and participatory which demands explanations of the design process by answering to open-ended questions;
2) models based on the GDPR's data protection impact assessments (DPIA), which obliges companies to perform a DPIA whenever data processing "is likely to result in a high risk to the rights and freedoms of natural persons."[167] Kaminski and Malgieri argue that the requirements of performing DPIA encompass, to some extent, AIA.[168] Some EU Member States, notably Slovenia, requires algorithmic impact assessments as a specific safeguard in case of automated decision-making under Article 22(1) of the GDPR;
3) a questionnaire model, such as the one proposed by the Canadian authorities which foresee the AIA as "a questionnaire designed to help you assess and mitigate the impacts associated with deploying an automated decision system." [169]

Finally, Selbst notes that a self-regulatory or ethics model of impact assessment and audits such as "social impact assessment" (SIA) or human rights impact assessments (HRIA), recommended by the UN Guiding Principles on Business and Human Rights, are yet another option.[170]

Importantly, AI HLEG notes that when implementing the above requirements, tensions may arise between them, which may lead to inevitable trade-offs. Any trade-offs should be explicitly acknowledged and evaluated in terms of their risk to ethical principles, including fundamental rights. Also, any decision about which trade-off to make should be reasoned and properly documented and should be continually reviewed. Some authors[171] suggest that the next steps

[165] Caplan and others (n 160).
[166] National Environmental Policy Act, 42 U.S.C. §§ 4331-47.
[167] Art. 35(1) of the GDPR.
[168] Kaminski M. E. and Malgieri G., 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' 29.
[169] https://canada-ca.github.io/aia-eia-js/.
[170] Selbst A., 'An Institutional View of Algorithmic Impact Assessments' 35 78.
[171] Whittlestone J. and others, 'The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions', *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society* (ACM 2019) <https://dl.acm.org/doi/10.1145/3306618.3314289> accessed 12 October 2020.

for AI ethics should indeed focus on bridging the gap between different sets of principles, acknowledge differences in values and identify ambiguities and knowledge gaps. The Guidelines propose that AI practitioners approach these ethical dilemmas and trade-offs 'via reasoned, evidence-based reflection'.[172]

The final component of the accountability requirements is redress, meaning that when unjust adverse impact occurs, accessible mechanisms should be foreseen that ensure adequate redress. This is key to ensure trust in AI.

In order to comply with the *accountability* requirement, the AI4EU research team encourages organizations to answer the following questions, based on the Assessment List for Trustworthy AI (Figure 9).

Legislation and Policy: Are there explicit legislation and/or other policies relevant to your system/resource?
a. The system is covered by an explicit clear, legal framework or sectorial formal policies, and we address these explicitly. (2)
b. We are aware of policy partially relevant to our system and address these sufficiently. (1)
c. We are not aware of any relevant legislation or policy and do not address these. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Right to contest/liability: Are users able to contest decisions/actions or demand human intervention?
a. Processes for contesting and/or demanding human intervention are set up and clearly available. (2)
b. Some contestability or intervention processes are available. (1)
c. It is not possible to contest the system's output nor to demand human intervention. (0)
d. We consider that these issues are not applicable to our case. (N/A)

Figure 9: The abbreviated assessment list: accountability requirement[173]

Conclusion

Requirements for Trustworthy AI are not a theoretical concept but should be "translated" into procedures and/or constraints on procedures, which should be anchored in the AI system's architecture. The Guidelines lists a series of technical[174] and non-technical[175] methods to ensure Trustworthy AI. The idea that compliance with norms can be implemented into the design of an

[172] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).
[173] Dignum and others (n 41).
[174] These include: Architectures for Trustworthy AI, Ethics and rule of law by design (X-by-design), Explanation methods, Testing and validating, Quality of Service Indicators.
[175] Such as Regulation, Codes of conduct, Standardisation, Certification, Accountability via governance frameworks, Education and awareness to foster an ethical mind-set, Stakeholder participation and social dialogue, Diversity and inclusive design teams.

AI system is key to implementing the requirements. Importantly, there is an interrelationship between the seven requirements: all are of equal importance, support each other, and should be implemented and evaluated throughout the AI system's lifecycle. Finally, the Guidelines point out that the realisation of Trustworthy AI is not a one-off exercise but a continuous process. Any changes to the implementation processes should occur on an ongoing basis.

3.2.2.2 EP Resolution 2020/2012(INL) on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and related Technologies

In October 2020, the European Parliament (EP) issued three Resolutions on the ethical and legal aspects of Artificial Intelligence software systems: i) Resolution 2020/2012(INL) on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and related Technologies, ii) Resolution 2020/2014(INL) on a Civil Liability Regime for Artificial Intelligence (section 3.2.4.3.), and iii) Resolution 2020/2015(INI) on Intellectual Property Rights for the development of Artificial Intelligence Technologies (section 3.2.3.1). Resolution 2020/2012(INL) on a Framework of Ethical Aspects of Artificial Intelligence, Robotics and related Technologies highlights the need for a human-centric and a human-created AI approach and the need to establish a risk-based approach to regulating AI.

Importantly in the AI4Media context, the Resolution acknowledges the growing potential of AI in the areas of information, media and online platforms, including as a tool to fight disinformation. However, if not regulated, the Resolution stresses, it might also have ethically adverse effects by exploiting biases in data and algorithms that may lead to disseminating disinformation and creating information bubbles. To this end, the Resolution emphasizes the importance of transparency and accountability of algorithms used by video-sharing platforms as well as streaming platforms, in order to ensure access to culturally and linguistically diverse content. Moreover, the EP notes that "whereas data analysis and AI increasingly impact on the information made accessible to citizens; whereas such technologies, if misused, may endanger fundamental rights to freedom of expression and information as well as media freedom and pluralism."[176]

The Resolution calls for a common Union regulatory framework for the development, deployment and use of AI, robotics and related technologies. The resolution stresses that such a regulatory framework on AI should be based on Union law and values and guided by the principles of transparency, explainability, fairness, accountability and responsibility. The EP notes that whereas the Union has a strict legal framework concerning the protection of personal data and privacy and non-discrimination, gender equality, environmental protection and consumers' rights which applies and will continue to apply in relation to AI, robotics and related technologies, certain adjustments of specific legal instruments may be necessary to address new challenges posed by the use of AI. To this end, the EP is concerned that the current Union legal

[176] 'Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies, European Parliament, P9_TA(2020)0275'.

framework may no longer be fit for the purpose of effectively tackling the risks created by AI, robotics and related technologies. The authors of the Resolution also point out that common ethical principles are only efficient where they are also enshrined in law, and those responsible for ensuring, assessing and monitoring compliance are identified. Ethical guidance, the Resolution continues, such as the principles adopted by the High-Level Expert Group on Artificial Intelligence, provides a good starting point but cannot ensure that developers, deployers and users act fairly and guarantee the effective protection of individuals.

For that reason, the EP calls on the EC for an effective, comprehensive and future-proof regulatory framework of Union law. The EP "expects the Commission to integrate a strong ethical approach into the legislative proposal requested in the annex to this resolution as a follow up to the White Paper on Artificial Intelligence (…)".[177] To that end, in the Annex to the Resolution, the EP attached a proposal for a Regulation on ethical principles for the development, deployment and use of AI, robotics and related technologies. The EP makes a set of concrete legislative proposals such as to apply a risk-based approach to AI regulation. It proposes a set of obligations for high-risk technologies such as full human oversight and control, safety, transparency and accountability of high-risk systems, the requirement of social responsibility and gender equality and environmental sustainability. It also provides that any software, algorithm or data used or produced by high-risk AI, robotics and related technologies developed, deployed or used in the Union shall be unbiased and not discriminate. Moreover, any natural or legal person shall have the right to seek redress for injury or harm caused by the development, deployment and use of high-risk AI. Finally, it also provides provisions on risk assessment, compliance assessment and a European certificate of ethical compliance. Some aspects of the EP's proposal for a Regulation have been considered by the EC in its AI Act proposal (see Section 4.1.3).

3.2.3 Intellectual property rights AI initiatives

With the recognition of many benefits and potential risks AI technologies could bring, the EC and the EP adopted different texts to harmonise and avoid fragmentations of the Intellectual Property (IP) framework in the Union, as well as fostering AI innovation in Europe. Thus, the EP first adopted a resolution on Intellectual Property Rights (IPR) for the development of AI Technologies in October 2020, later followed by an action plan on IP adopted by the Commission in November 2020. The following sections first give a brief overview of the Resolution's objectives and recommendations, then elaborate on the action plan on IP. Finally, it provides a comprehensive analysis on the current state of art concerning IPR in the Union, with the aid of a study conducted for the Commission by University of Amsterdam's Institute for Information Law and the Joint Institute for Innovation Policy.

[177] ibid.

3.2.3.1 EP Resolution 2020/2015(INI) on Intellectual Property Rights for the development of Artificial Intelligence Technologies

As already mentioned, the second resolution adopted by the EP in October 2020 was the "European Parliament resolution on Intellectual Property Rights for the Development of Artificial Intelligence Technologies (2020/2015(INI))".[178] According to the Resolution, the Union seeks not to fall behind in the global AI competition and become the world leader in AI technologies instead, while regaining and safeguarding the Union's digital and industrial sovereignty, ensuring its competitiveness, and promoting and project innovation. Thus, to achieve this aim an effective Intellectual Property system is required. Accordingly, the system shall be fit for the digital age, enabling innovation, ensuring strong economic growth and citizens' prosperity, while following a human-centered approach, compliant with ethical principles and human rights. Therefore, the EP Resolution calls for the following objectives to be met, shown in Figure 10 below.

[178] European Parliament, the Resolution of 20 October 2020, on "Intellectual Property Rights for the Development of Artificial Intelligence Technologies (2020/2015(INI)." https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.pdf.

Figure 10: EP (IP) Resolution's Objective and Recommendations[179]

Conclusively, the Resolution lays out its recommendations specifically tailored for different branches of IP such as copyright, patent, and database protection.

Copyright

The EP suggests that the EC should support a horizontal, evidence-based and technologically neutral approach to common, uniform copyright provisions applicable to AI-generated and AI-assisted works in the Union. The difference between AI-assisted human creations and AI-

[179] European Parliament, European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies. Available at: https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.pdf.

generated creations should be distinguished. The latter poses challenges concerning ownership, inventorship, appropriate remuneration, and issues related to potential market concentration.

Autonomously AI-Generated Creations

Under the current EU acquis, work autonomously created by AI and robots might not be eligible for copyright protection, as they will not be able to fulfil the originality requirement, which is linked to a natural person and author's intellectual creation. However, if such work is considered to be copyrightable, it is recommended that ownership of rights should only be assigned to natural or legal persons that created the work lawfully and only if authorization has been granted by the copyright holder if copyright protected material is being used, unless copyright exceptions and limitations apply.

Ownership of autonomous AI-generated output with an artistic nature would not be appropriate, as it creates a negative impact regarding incentives for human creators if AI technologies were to be granted legal personality.

AI-Assisted Creations

Where AI is used only as a tool to assist an author in the process of creation, the current IP framework remains applicable. However, the EP emphasized that more thorough research is necessary for the purpose of evaluating human input regarding AI algorithmic data. In this research, the priority should be given to assessment by sector and type of IPR implications of AI technologies. This approach should take into account the degree of human intervention, the autonomy of AI, the importance of the role and the origin of the data and copyright-protected material used and the involvement of other relevant factors.

The Resolution also emphasizes that the legal challenges of reverse engineering, which is an exception to the copyright protection of computer programs and the protection of trade secrets, should be taken into account in the context of development of AI technologies, as they pose a crucial importance for innovation and research.

Patent

The EP adopts a pro-innovation point of view by highlighting that technical creations generated by AI technology must be protected under the IPR legal framework, in order to encourage investment in this form of creation and improve legal certainty for citizens, businesses, and inventors. Thus, as explained below in detail, the report touches upon patent protection, establishment of industry, and the patent protection framework's prominence for innovation.

Protection

First, patent protection can be granted provided the invention meets the patentability test. Second, mathematical methods are excluded from patentability unless they are used for a technical purpose in the context of technical inventions, which are themselves patentable only

if the applicable criteria relating to inventions are met. Third, if an invention relates either to a method involving technical means or to a technical device, it is considered as a whole and considered technical in nature, enabling the invention not being excluded from patentability.

INCENTIVIZING INNOVATION

The Resolution takes a good note of the role of the patent protection framework in incentivizing AI inventions and promoting their dissemination, as this would create opportunities for European companies and start-ups and foster the development and uptake of AI in Europe. Standard essential patents play a key role in the development and dissemination of new AI and related technologies and in ensuring interoperability, therefore, the Commission's support regarding the establishment of industry standards and encouragement of formal standardization gains prominence.

Database Protection

The Parliament emphasizes that the European Data Strategy must ensure a balance between promoting the flow of wider access to and the use of and sharing of data on the one hand, and the protection of IPRs and trade secrets on the other, while respecting privacy and data protection rules. To sum up:

- Comprehensive information should be provided on the use of data protected by IPRs, in particular in the context of platform-to-business relationships. Therefore, the Commission's intention to create a single European data space is welcomed by the Resolution.
- Further clarification is needed regarding the protection of data under copyright law and trademark and industrial design protection for autonomous AI-generated work.
- The lawful use of copyrighted works, other subject matter, associated data, including pre-existing content, high-quality datasets, and metadata needs to be assessed in the light on the existing rules on limitations and exceptions to copyright protection, i.e., the text and data mining exception, as provided for by the Directive on copyright and related rights in the Digital Single Market.
- Facilitating access to data and data sharing is prominent, as well as doing the same for open standards and open technology, to encourage investment and boost innovation.
- IPR issues arising from the creation of deep fakes on the basis of misleading, manipulated or simply low-quality data, irrespective of such deep fakes containing data which may be subject to copyright should be further clarified.

Use of Non-Personal Data by AI Technologies

The Resolution also gives a substantial place to the use of non-personal data by AI technologies, by highlighting the following:

(i) Full implementation of the Digital Single Market Strategy is prominent to improve the accessibility and interoperability of non-personal data in the EU.

(ii) AI generated output must not be discriminatory and that one of the most efficient ways of reducing bias in AI systems is to ensure, to the extent possible under Union law, that the maximum amount of non-personal data is available for training purposes and machine learning. Therefore, the EP suggests that the Commission should reflect on the use of public domain data for such purposes.
(iii) Voluntary non-personal data sharing between businesses and sectors should be promoted and based on fair contractual agreements, including licensing agreements.
(iv) In order to solve the issues concerning relationships between economic operators whose purpose is to make use of non-personal data, the Resolution endorses a possible revision of the Database directive and a possible clarification of the application of the directive on the protection of trade secrets as a generic framework.

3.2.3.2 Action Plan on Intellectual Property

In November 2020, the EC adopted an action plan on IP titled, "Making the most of the EU's innovative potential: An intellectual property action plan to support the EU's recovery and resilience."[180] With this plan, the EC recognized the need to upgrade the system for IP protection in data economy and society in the face of technological revolution. According to the plan, new technologies, such as AI, can facilitate the protection of IP, improve transparency, allow for a smoother distribution of license fees, and more effectively tackle counterfeiting and piracy. To explore the full potential of these new technologies and make the most out of the abovementioned benefits, the EC proposed to take the following actions:

- Encouraging an industry dialogue;
- Engaging in stakeholder discussions;
- Mapping and analysing all issues, including the clarification of the ownership and authorship status of AI-assisted output and AI-generated output; and
- Addressing harmonisation gaps to avoid fragmentation in the IP framework in the EU.

3.2.3.3 The Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework Report

In order to better depict the EP Resolution, the Action Plan on IP, and current state of art concerning the IPR, i.e., copyright, patent, and sui generis database rights, this chapter will be analysed in conjunction with "The Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework Report" prepared by IViR (University of Amsterdam Institute for Information Law) & JIIP (The Joint Institute for Innovation Policy) for

[180] European Commission, Communication of 25 November 2020, "Making the most of the EU's innovative potential, An intellectual property action plan to support the EU's recovery and resilience." https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0760.

the EC under the study reference of SMART 2018/0052.[181] The study will be hereinafter referred to as 'the Report.'

Current State of Art on Copyrightability

The EU acquis expressly harmonizes four categories of copyright-protected subject matter: computer programmes, databases, photographs, and works of visual art.[182] According to the CJEU case-law[183], such works are protected because they are "the author's own intellectual creation." Therefore, under the current EU copyright law, in order for a creation (human or AI-created) to qualify as a (protected) work, the four-step test depicted below (Figure 11) must be met.

Figure 11: Copyrightability[184]

Accordingly, tailored analysis of the four-step-test concerning AI-assisted output, borrowed from the Report's language, is presented below in detail (Table 2).

Production in literary, scientific or artistic domain	The Berne Convention states, to be considered a "work," a creation must be produced within the "literary, scientific, or artistic domain."[185] However, it is not clear whether this translates into a substantive requirement under EU copyright law. Assuming it does, many AI-assisted output would be able to pass the first step of the four-step-test, as these systems are capable of generating most of the work mentioned in Art. 2(1) of the Berne Convention.

[181] Directorate-General for Communications Networks, Content and Technology (European Commission), and others, "Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework : Final Report." Publications Office of the European Union, 2020. https://data.europa.eu/doi/10.2759/458120.
[182] Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs Art 1(3); Directive 96/9/EC, Art. 3(1) Database Directive; Directive 2006/116/EC, Art. 6 Term Directive; Directive 2019/790, Art. 14 Copyright on Digital Single Market Directive (on works of visual art in the public domain).
[183] CJEU C-05/08 Infopaq International v Danske Dagblades Forening (2009) ECLI:EU:C:2009:465 (Infopaq), CJEU C-310/17 Levola Hengelo BV v Smilde Foods BV (2018) ECLI:EU:C:2018:899; CJEU C-469/17 Funke Medien NRW GmbH v Bundesrepublik Deutschland (2019) ECLI:EU:C:2019:623 (Funke Medien); CJEU C-683/17 Cofemel – Sociedade de Vestuário SA v G-Star Raw CV (2019) ECLI:EU:C:2019:721 (Cofemel).
[184] Berne Convention, 1886, Art. 2 – Art. 13.
[185] Berne Convention, 1886, Art. 2.

Human Intellectual Effort	Along with the Berne Convention, the CJEU case-law emphasizes the requirements of a work reflecting a "creative choice" of a human being.[186] Therefore, personal touch of a natural person is deemed to be required, in order to pass this step. The criterion of human intellectual effort excludes from copyright protection of work that is produced without any human intervention, i.e., a work created by an autonomous robot without any human intervention. However, this requirement does not rule out AI-assisted output, as it is plausible to create works of authorship with the aid of an AI technology. This includes works where there is little human intervention in the AI-assisted output. Because for the foreseeable future, according to the Report, it is hard to expect an AI-generated creative work without involvement of any natural person.[187]
Originality/Creativity	Originality criterion is a twofold requirement; first, the subject matter must be "author's own,'' and second, it must constitute an "intellectual creation" by the author. Nonetheless, originality does not mean that an artistic merit or aesthetic quality is expected to provide protection for works. Additionally, the Report states, what EU copyright law focuses on is the act of creation in terms of making free and creative choices. For that reason, "for an AI-assisted output to pass the test of originality/creativity, it is sufficient that the output be the result of creative choices. These choices may occur at several stages of the creative process: conception, execution, and/or finalization." Concerning the minimum level of originality, though the CJEU precedent differs in some ways, the Report suggests that even a combination of fairly obvious choices in the design, execution, and editing of an AI-assisted output could suffice. Thus, mere human intervention at the conception and redaction stages could suffice for copyright protection.
Expression	This criterion implies a "causal link" between author's creative act (the exercising of their creative freedom) and the expression thereof in the form of the work produced. The Report suggests that “the concept of a work as the author's own intellectual creation not merely requires human agency or intervention, but also some degree of authorial intent.”[188] Hence, the Report adds, as long as the output

[186] CJEU Case C-145/10 Eva-Maria Painer v Standard VerlagsGmbH and Others 2013 ECLI:EU:C:2011:798; CJEU C-683/17 Cofemel – Sociedade de Vestuário SA v G-Star Raw CV (2019) ECLI:EU:C:2019:721 (Cofemel)c.

[187] Ginsburg J., “The Concept of Authorship in Comparative Copyright Law,” 2014: "The participation of a machine or device, such as a camera or a computer, in the creation of a work need not deprive its creator of authorship status, but the greater the machine’s role in the work’s production, the more the “author” must show how her role determined the work’s form and content.”

[188] Burk, Dan L., “Thirty-Six Views of Copyright Authorship”, Houston Law Review, Vol. 58, 2020. https://ssrn.com/abstract=3570225.

	"reflects creative choices by a human being at any stage of the production process, an AI-assisted output is likely to qualify for copyright protection."

Table 2: Copyrightability of AI Assisted Output Steps[189]

It is important to note that not every AI-assisted output qualifies as copyright-protected works. For instance, this may involve mundane AI-assisted output like weather forecasts or news reports if they leave only limited space for creative choices by a natural person.

Authorship

Though the CJEU on various occasions[190] suggested that the notion of "author" is reserved for a human creator, rules on authorship and copyright ownership are largely unharmonized in the EU, as the discretion to decide on the authorship and ownership status of a work is left to national laws and courts of the Member States. Therefore, without going into detail on Member States' diverging legislation, it is only possible to give a general overview on the state of art of authorship in the EU. First, the Report suggests, if AI-assisted output does not qualify as a protected work, no authorship can exist. Such authorless production might however still enjoy protection under related rights. Second, in the cases where there is more than one author involved in the process, individually or collectively engaging in the creative choices, co-authorship could exist. This is possible even if the creative contributions occur at different stages of the creative process (i.e., conception, execution, and redaction.) Third, the rules on whether the authorship should be granted to the user or the developer of the AI system might not be clear in some specific situations. In order to analyse the authorship status, it should be looked into whether one of them or both exercised free choices at any stage of the creative process. If the developer exercised free choices, while the role of the user in the system was constricted, he/she will qualify as the sole author of the protected work. The exact opposite scenario, as well as co-authorship, is also possible here.

Ownership

Traditionally, ownership follows authorship, as copyright vests in the person having created the work by default. On the other hand, it is important to note that there are some exceptions and diverging approaches at the national level to this rule (i.e., some national laws and the Berne

[189] The Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework Report. Available at: https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights.

[190] CJEU Case C-277/10 Martin Luksan v Petrus van der Let 2012 ECLI:EU:C:2012:65; CJEU Case C-572/13 Hewlett-Packard Belgium SPRL v Reprobel SCRL ECLI:EU:C:2015:750.

Convention[191] provide for legal presumptions of authorship/ownership in favour of the person "whose name appears on the work in the usual manner"[192]).

Protection by Related Rights

As the Resolution does not touch upon these rights much, they will only be explained briefly here. Related rights, also known as neighbouring rights, protect the legal interests of persons or legal entities, when their subject matters do not qualify as works under copyright law.[193] The major difference between copyright and related rights is that unlike copyright, related rights do not require originality or authorship. Currently, under EU Law, the following six related rights are recognized (Table 3):

Rights of Phonogram Producers	Variety of AI produced audio output could be protected under this related right.
Rights of Broadcasters	Automatically produced and transmitted broadcasts by AI systems could qualify for this protection.
Rights of Film Producers	As this right does not require originality or provide for any other threshold requirement, it allows protection of all sorts of video content generated by AI systems, varying from surveillance videos, to drone footage, to satellite imagery, to video content automatically generated for media channels.
Rights of Publishers of Press Publications	A new related right brought by the Directive on Copyright in the Digital Single Market Art 15 and Art 26. Since there is no originality requirement under this right, content generated by AI applications such as a blog post generated by AI or publishing sports news would probably qualify for protection.
Other Related Rights in National Laws	Member States have a diverse mix of related rights. For the sake of relevance, this part will not be analyzed.
Sui Generis Database Right	This right will be discussed in detail later in the chapter. However, unlike the other five rights listed above, sui generis database right protection requires for a creation to meet a certain threshold.

Table 3: Neighbouring Rights in the EU[194]

[191] Berne Convention, 1886, Article 15.
[192] Ginsburg J., The Concept of Authorship in Comparative Copyright Law, 52 DePaul L. Rev. 1063, 2003. https://via.library.depaul.edu/law-review/vol52/iss4/3
[193] World Intellectual Property Organization, *Understanding Copyright and Related Rights*. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.
[194] The Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework Report. Available at: https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights

Data and Database Protection

COPYRIGHT PROTECTION

A database that is an original intellectual creation, in other words, satisfying the copyrightability requirements mentioned above, could be protected through copyright. Importantly, this protection guarantees to protect the structure of the database and not its content.[195]

SUI GENERIS PROTECTION

The Database Directive provides a sui generis protection to a database that is "a collection of independent works, data, or other materials arranged in a systematic or methodical way and individually accessible by electronic or other means."[196] Unlike copyright protection to databases, this right includes the protection of the content of a database, which the maker can prevent the extraction and/or reuse of the whole or a substantial part of the database's content. The Directive also requires a substantial investment, either in a "qualitative" and/or a "quantitative" way. Therefore, it differs from other related rights discussed above, as they do not require for a creation to meet a certain threshold. Furthermore, the Report suggests that most databases in practice will result from quantitative investment, involving "deployment of financial resources and/or the expanding of time, effort, and energy."[197]

Another interesting fact about this protection is that the EU database right does not require human authorship, it allows for protection of all sorts of AI-assisted outputs that qualify as databases, including weather reports and sports data generated by AI (which are not copyright-protected work as they do not meet the 4-step test as discussed above.) Nevertheless, according to the CJEU case-law in order to qualify for the protection, the database must be "arranged in a systematic or methodical way". [198] According to the author of the Report, this means the raw machine-generated data is left out of the scope of the sui generis database protection.

Patentability

The European Patent Convention (EPC) makes a distinction between a procedural right to the patent under Art. 60, which is deemed to belong to the patent applicant, and the substantive right to the patent, as the right to a European patent shall belong to the inventor or his successor in his title. However, the Report notes that the relation between the applicant and the person

[195] Europa, Database Protection. https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm.
[196] Directive 96/9/EC, Database Directive, Article 1(2).
[197] Directive 96/9/EC, Database Directive, Recital 40.
[198] CJEU C-46/02 Fixtures Marketing Ltd v Oy Veikkaus Ab (2004) ECLI:EU:C:2004:694 (Fixtures Marketing Ltd v Oy Veikkaus Ab); CJEU C-203/02 The British Horseracing Board Ltd and Others v William Hill Organisation Ltd. (2004) ECLI:EU:C:2004:695 (British Horseracing Board and others); CJEU C-338/02 Fixtures Marketing Ltd v Svenska AB (2004) ECLI:EU:C:2004:696 (Fixtures Marketing Ltd v Svenska AB); CJEU C-444/02 Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP), (2004) ECLI:EU:C:2004:697 (Fixtures Marketing Ltd v OPAP).

having the substantive right is governed by national laws.[199] Therefore, the European Patent Office (EPO)[200] has no power to determine disputes regarding substantive entitlement.

Patentability of AI-assisted Output

For an invention to be protected under patent law, the four-step-test demonstrated below must be met (Figure 12).

Figure 12: Patentability [201]

It is important to note that the first two steps of the test do not concern patentability of AI-assisted output as they usually belong to a (or any) field of technology, as well as being industrially applicable. Therefore, the analysis in Table 4 below only focuses on Steps 3 and 4.

Novelty	The EPC Art. 54 defines novelty as, "An invention shall be considered to be new if it does not form part of the state of the art… The state of the art shall be held to comprise everything made available to the public by means of a written or oral description, by use, or in any other way, before the date of filing of the European patent application." To defeat novelty, the EPO guidelines suggests "if the information given therein is sufficient to enable the skilled person, at the relevant date of the document, to practice the technical teaching which is the subject of the document, taking into account also the general knowledge at that time in the field."[202] As the Report suggests, whether AI is involved or not, determining novelty can always be a difficult process.[203] Fortunately, as outlined below, the role of AI in this field means that several parallel changes are happening. Because the issue is not only patentability of AI-assisted output, but also AI applications' use in assessing novelty. *Quantitative Changes:*

[199] Visser and others, Visser's Annotated European Patent Convention (EPC), 138.
[200] European Patent Convention (EPC), 1973, Article 60. https://www.epo.org/law-practice/legal-texts/html/epc/2016/e/ar60.html.
[201] EPO, Guidelines for Examination, Patentability Requirements. https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_i_1.htm.
[202] EPO, Guidelines for Examination, G-VI, 4. Enabling disclosure of a prior-art document. https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vi_4.htm.
[203] ibid.

	More data can be parsed by AI systems used by patent applicants or patent offices. AI can assist humans in selecting the most relevant data to work with. AI systems can also be used by patent offices to analyse more potentially relevant prior art faster. *Qualitative Changes:* Regarding possible qualitative changes, the novelty assessment would be increasingly performed by AI systems, rather than pure human review.
Inventiveness	According Art. 56 of the EPC, an invention is considered to involve an inventive step if, having regard to the state of the art, it is not obvious to the person having ordinary skills in art (POSITA). [204] There are three main steps in the inventiveness analysis: *Step 1: Determining the closest prior art;* *Step 2: Establishing the objective technical problem; and* *Step 3: Considering whether or not the claimed invention, starting from the closest prior art and the objective technical problem, would have been obvious to the skilled person.*[205] According to the author of the Report, AI may change the three steps in the following ways: - The use of AI has already changed the step 1 process, which is typically based on a fairly straightforward data analysis, namely identifying features and finding the closest match in the dataset. Beyond a certain level of complexity of the claimed AI-assisted invention, EPO examiners would have difficulty in establishing the causal link required to a finding of obviousness. Therefore, it is hard for applications concerning AI-assisted outputs to be rejected on the grounds of lacking the inventive step. - Determination of obviousness should still be conducted by a human, as Step 2 and Step 3 of the assessment require cognitive functions that only humans possess at the moment. - AI systems could possibly change the analysis of the POSITA, as the use of AI innovation could lead to AI processing more data much faster than a human applicant or examiner and finding correlations that a human may not find.

Table 4: Patentability of AI Assisted Output[206]

[204] EPC Article 56.
[205] European Patent Office, Guidelines for Examination, G-VII (5). https://www.epo.org/law-practice/legal-texts/html/guidelines/e/g_vii_5.htm.
[206] The Trends and Developments in Artificial Intelligence - Challenges to the Intellectual Property Rights Framework Report. Available at: https://ec.europa.eu/digital-single-market/en/news/trends-and-developments-artificial-intelligence-challenges-intellectual-property-rights.

Inventorship

Though main international treaties seem to imply that an inventor is a natural person, these instruments predate the emergence of AI technologies and AI-assisted or generated output. As at this point in time, outputs that are autonomously generated by AI do not exist, this analysis excludes them. The Report implies that the key question and confusion concerning inventorship is whether "human inventorship" is a substantive patentability requirement or rather merely a requirement. Fortunately, the EPO clarifies this confusion by stating that an AI system cannot be the named inventor on a patent application.[207] Therefore, as the Report's author suggests, naming the inventor is deemed to be a formal requirement in a way that a human person be named as inventor. Finally, as noted above, the specific issue of inventorship is at the discretion of national courts of the Member States, not at the level of the EPO and the EPC.

Ownership

In terms of the owner of an invention, there are three possible claimants: "(1) The programmer or developer of the AI system; (2) The owner of the AI system; and (3) The (authorized) user of the AI system who provided with training or other data." Unfortunately, there are no clear-cut rules on ownership as neither the TRIPS nor the Paris Convention provides clear rules concerning ownership of patents. However, as the Report suggests borrowing from EPC's terminology, "AI systems do not have legal personality and, therefore, cannot be "employees or have a successor in title in accordance with the law of the State in which the employee is mainly employed." In sum, the question of ownership, similar to inventorship, cannot be addressed with a single answer, as these issues mostly concern national law and courts of the Member States.

Disclosure Requirement

Regarding disclosure, Art. 83 of the EPC states, "The European patent application shall disclose the invention in a manner sufficiently clear and complete for it to be carried out by a POSITA.''[208] The aim behind this requirement is to ensure reproducibility and plausibility of the claimed invention. Nonetheless, the requirement might complicate things with the use of AI systems in the inventive activity, "as black box nature of certain AI systems may make it challenging to provide a sufficiently clear and complete disclosure for the invention to be carried out by a POSITA." Therefore, like inventiveness, this is a matter that should likely to be solved primarily by patent offices. Though, as a matter of recommendation, the author of the Report suggests, "the patent application must disclose enough for replicability by a POSITA, but not more than is the case for non-AI assisted inventions."[209]

[207] EPO decision of 27 January 2020, EP 18 275 163. https://register.epo.org/application?number=EP18275163#_blank.
[208] EPC Art. 83. https://www.epo.org/law-practice/legal-texts/html/epc/2020/e/ar83.html.
[209] Directorate-General for Communications Networks, Content and Technology (European Commission), and others, "Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework : Final Report." Publications Office of the European Union, 2020. p. 113. https://data.europa.eu/doi/10.2759/458120.

3.2.4 Safety and Liability AI Initiatives

AI in many of its aspects comes with promises and risks; it goes the same way for safety and liability. Sufficient safeguards are needed to minimise the risks of harm. In this introduction, we will first briefly present some EU initiatives prior to 2019. Then, we will analyse more recent policy initiatives including: the EC expert group "Report on liability for Artificial Intelligence and other emerging technologies", "the Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics," and lastly, the "European Parliament resolution on automated decision-making processes: ensuring consumer protection and free movement of goods and services".

The EP started in 2015 by setting up a working group with the primary aim of drawing up "European" civil law rules on emerging technologies and robotics. This working group delivered a draft report setting out a series of recommendations on civil law rules on robotics.[210] In 2016, the EP continued its initiatives by commissioning a study on European Civil Law Rules for Robotics.[211] The study assessed the main challenges that emerging technologies raise for the civil law landscape. Following this analysis, the EP adopted in 2017 a resolution with recommendations to the Commission on Civil Law Rules on Robotics.[212] A year later, the EPRS delivered a study on 'a common EU approach to liability rules and insurance for connected and autonomous vehicles – European added value assessment'.[213] In the aftermath, the EC set up an expert group on liability and new technologies.[214] The expert group report on liability will be the first point of analysis of this section.

3.2.4.1 Report on Liability for Artificial Intelligence and other emerging technologies

On 27th November 2019, the expert group on Liability and New technologies released its report on liability for AI and other emerging technologies.[215] The report investigates the civil liability

[210] European Parliament, 'Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect.

[211] European Parliament, 'Study for the JURI Committee on European Civil Law Rules for Robotics', (2016), https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf.

[212] European Parliament, 'Resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics' (2015/2103(INL)), https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52017IP0051&from=EN.

[213] European Parliamentary Research Service, Study on a common EU approach to liability rules and insurance for connected and autonomous vehicles (2018), https://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf.

[214] European Commission, 'Expert Group on liability and new technologies', https://ec.europa.eu/transparency/expert-groups-register/screen/expert-groups/consult?do=groupDetail.groupDetail&groupID=3592&NewSearch=1&NewSearch=1.

[215] European Commission Expert Group on Liability and New Technologies, 'Report on liability for Artificial Intelligence and other emerging technologies' (2019), https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en.

challenges raised by digital technologies and puts forward recommendation on how to adapt the current legal framework on liability. The report provides a state-of-the-art analysis of the existing laws in Europe which deals with liability for emerging technologies. The report does not go in detail of the sectoral specific legislation applicable to liability which can be numerous but chooses to stay on a macro-level analysis. Besides some harmonized legislations' liability regimes still vary greatly from one Member State to another, therefore, some might think that the report does not go sufficiently in depth and further research needs to be conducted on this aspect.[216] The Product Liability Directive is a cornerstone of the EU harmonisation effort on liability, it is based on the principle that the producer (broadly defined along the distribution channel) is liable for "damage caused by the defect in a product they have put into circulation for economic purposes or in the course of their business"[217]. Even if the Directive was drafted in a technological neutral way, the report points out that some key elements are today inadequate for addressing the potential risks of emerging digital technologies. [218] This includes the scope of the directive and the notion of product and defect. There are also new procedural challenges associated with emerging technologies.[219] In a second part, the report points the key concepts underpinning classical liability regimes which would need legal clarification given the emerging technologies specificities. It also establishes and expand on new specific rules, principles and concept which might be necessary to adopt.[220]

The report points out that only basic protection of victim is ensured through the liability regimes in place in Member States, in case damage is caused by an emerging technology. The traditional liability rules are not a best fit to meet the challenges raised by the emerging technologies. Most of the time liability rules are part of a legal corpus written decades or century ago. As an illustration the notion of damage, causal link and fault required in tort law can become tricky proof in a case involving emerging technologies.

For these reasons, the report provides further recommendations, certain measures and approaches to improve the current liability regimes. These are summarised below (Table 5).

Operator's strict liability	Strict liability (without fault) should be applicable primarily to emerging digital technologies operating in public spaces that may

[216] Dheu Orian, 'EU report on AI, new technologies and liability : key take-aways and limitations' (2020), https://www.law.kuleuven.be/citip/blog/eu-report-on-ai-new-technologies-and-liability-key-take-aways-and-limitations/.
[217] EC Expert Group on Liability and New Technologies Report, p. 27.
[218] European Commission, Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC), COM(2018) 246 final, https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2018)246&lang=en.
[219] EC Expert Group on Liability and New Technologies, op. cit., Report p.29.
[220] ibid, p.32.

	typically cause significant harm with victim's easy access for compensation.[221]
Competing control of providers	Firstly, the experts put forward that the concept of 'operator' is a more neutral and flexible concept that owner, user/keeper when it comes to emerging technologies.[222] Whenever competing control over the emerging technology occurs, liability should lie with the one who has more control over the risks of operation.
Data logging	Emerging digital technologies enable the identification inter alia of what has caused an accident and this can be achieved through log files. Therefore, the report points out that appropriate logging requirements should be defined and failure to comply with a logging and disclosure duty should lead to a rebuttable presumption of liability. [223]
Safety rules	Experts suggest that failure to comply with the safety rules that would have prevented harm to occur should lead to the reversal of the burden of proof. This would incentivise the compliance with safety rules developed by lawmakers.[224]
Burden of proof	The report outlines that as a general rule, the victim should continue to be required to prove what caused her harm. However, reversal should be granted especially where it becomes unreasonably difficult for the victim to prove the constitutive elements given the emerging technologies which caused this harm. [225]
Damage to data	Liability should arise through contractual liability, for instance where there was an intention to cause harm to the data by the deletion, deterioration, contamination, encryption, alteration, or suppression of data.[226]
Legal personality	There is no need to give a legal personality to emerging digital technologies in accordance with expert's opinion on the subject. Indeed, the report resists to the temptation to fall in science fiction and underlined that harm caused are reducible to risks "attributable to natural persons or existing categories of legal persons, and where this is not the case, new and specific laws

[221] EC Expert Group on Liability and New Technologies, Report op. cit., p.39
[222] ibid, p.41.
[223] ibid. p.47-48.
[224] ibid, p.49.
[225] ibid, p.49.
[226] ibid, p.59.

	directed at individuals are a better response than creating a new category of legal person".[227]

Table 5: Recommendations on liability regime for AI

The report, while putting forward avenues of reflection and identifying the main challenges, does not provide for concrete solutions or recommendations on how and by whom these legal regimes be amended. [228] However, it acknowledges that the one-size-fits-all approach is not compatible with the wide variety of liability regime and the complexity of the diverse emerging technologies.

3.2.4.2 Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics

The report from the EC on the safety and liability implications of AI, the Internet of Things and robotics[229] was released in February 2020. The report accompanied the White Paper on AI presented earlier in Section 3.2.1.3.

Similarly, to previous studies and reports, it assesses the relevant legal framework, and identifies the challenges and uncertainties of the liability, but not only those, as it also includes the safety framework. Indeed, the two frameworks are complementary mechanisms pursuing the same goal of ensuring a functioning and safe internal market for products and services. The report indicates that the current product safety and liability legislation already supports an extended concept of safety protecting against all kind of risks arising from the product according to its use. The wheel does not need to be fully reinvented, however, numerous adaptations and new provisions covering not-yet-addressed risks by the current legal framework are necessary to provide legal certainty.

Safety

The "Union product safety legislation aims to ensure that products placed on the Union market meet high health, safety and environmental requirements and that such products can circulate

[227] ibid, p. 38 ; Abbott, Ryan Benjamin and Sarch, Alex F., Punishing Artificial Intelligence: Legal Fiction or Science Fiction (February 1, 2019). 53 UC Davis Law Review 1, 323 (2019), Available at SSRN: https://ssrn.com/abstract=3327485 or http://dx.doi.org/10.2139/ssrn.3327485.

[228] Orian D., 'EU report on AI, new technologies and liability : key take-aways and limitations' (2020), https://www.law.kuleuven.be/citip/blog/eu-report-on-ai-new-technologies-and-liability-key-take-aways-and-limitations/.

[229] European Commission, Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC), COM(2018) 246 final, https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52020DC0064.

freely throughout the Union"[230]. The report concludes that the product safety legislation is adequate as it already benefits from a flexible scope of protection for the risks arising for products but that new provisions should be introduced to complement the framework to bring more legal certainty.

The report analysed the horizontal rules providing the coherent basis for the sectorial rules on product safety including: (i) General Product Safety Directive[231]; (ii) The New approach framework [232]; (ii) The Market Surveillance Regulation. [233] The report also highlighted the key role of European Standardisation in the Union product safety legislation.

Challenges of AI systems include complex value chains, complex products, services, and systems, opacity, data dependency, connectivity, risks for mental health of users, autonomy and self-learning features.

To address these challenges the following measures were put forward. A new risks assessment procedure could be put in place before the product or service enters the market when the product is subject to important changes during its lifetime in combination with additional instruction and warning for the users.[234] Because of their autonomy and self-learning aspects, the report puts forward that for avoiding cases where decisions deviate from what was initially intended, human oversight through the whole life cycle of the AI product and systems is recommended.[235] Explicit obligations for producers in respect of mental safety risks to users, for instance in case of collaboration or interaction with humanoid robots, should be envisaged. [236] The report raises the question as to whether the Union product safety legislation should provide for specific requirements addressing the risks to safety of faulty data at the design stage as well as mechanisms to ensure that quality of data is maintained throughout the use of the AI

[230] European Commission White Paper on AI, p. 4.
[231] Directive 2001/95/EC the European Parliament and of the Council of 3 December 2001 on general product safety, OJ L 11, 15.1.2002, p. 4–17.
[232] Regulation (EC) No. 2008/765 setting out the requirements for accreditation and market surveillance relating to the marketing of products and Decision (EC) No. 2008/768 on a common framework for the marketing of products, OJ L218, 13.8.2008, p. 30–47.
[233] Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93, OJ L 218, 13.8.2008, p. 30–47, ELI: http://data.europa.eu/eli/reg/2008/765/oj, and, from 2021 onwards, Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/201,1 OJ L 169, 25.6.2019, p. 1–44, ELI: http://data.europa.eu/eli/reg/2019/1020/oj.
[234] European Commission report to the European Parliament, the Council and the European Economic and Social Committee, on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, 19 February 2020, OJ, C(2020) 64 final, p. 6-7.
[235] ibid. p. 7-8.
[236] ibid., p. 8.

products and systems.[237] The current framework does not provide adequate provisions to solve the risks derived from the opacity of the systems. Therefore, the report outlines that the requirements for transparency of algorithms, as well as for robustness, accountability and when relevant, human oversight and unbiased outcomes should be established in combination with an ex-post mechanism of enforcement. This will be key to build trust in the use of those technologies.[238] Existing rules should be adapted and clarified when it comes to stand-alone software solutions.[239] The current system of shared responsibility and complex value chains as present in the current legal framework needs adaptations given the particularities of the AI value chain. Legal certainty can be brought with provisions framing the cooperation between the economic actors in the supply chain.[240]

Liability

Emerging technologies and AI systems own many characteristics which challenge the traditional liability concepts and legal mechanisms. This creates uncertainties for victims, and the report insists that while safeguarding innovation, persons having suffered harm caused with the involvement of AI systems need to enjoy the same level of protection as persons having suffered harm caused by other technologies. The scope of the Product Liability Directive and the notion of putting into circulation should also be further clarified to reflect better the characteristics of AI systems and ensure legal certainty for the economic actors.[241] Some open questions remain in the report such as whether and to what extent it may be needed to mitigate the consequences of complexity by adapting the burden of proof required by national liability rules for damage caused by the operation of AI applications.[242] Another point of interrogation is whether the extension of strict liability for high risks AI systems such as operating motor vehicles, airplanes or nuclear power plants or whether coupling strict liability with a possible obligation to conclude available insurance would be suitable to compensate victim's damage.[243] Concerning the burden of proof for causation and faults, the EC also reflects whether it needs to be adapted.

Based on this report, the White Paper on AI drew the conclusion that beside the adjustments to the existing legislation, a new and specific legislation specifically for high risks AI-systems may be needed to make the EU legal framework fit for AI. [244] The White Paper is more moderated than the report we just analysed. The White Paper acknowledges that the allocation of responsibilities between the different actors must be improved, the fault-based liability schemes

[237] ibid, p. 9.
[238] ibid.
[239] Ibid, p. 11.
[240] Ibid.
[241] ibid., p. 12, p. 15.
[242] ibid., p. 14.
[243] European Commission report to the European Parliament, the Council and the European Economic and Social Committee, on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, 19 February 2020, OJ, C(2020) 64 final, p. 16.
[244] European Commission White Paper on AI, p. 4.

might not be adapted for the AI systems, the level of protection for individuals must be guaranteed while caused by AI-systems or not. The document also outlines the importance of adopting a common approach at the EU level.

3.2.4.3 European Parliament recent initiatives

Resolution on automated decision-making processes: ensuring consumer protection and free movement of goods and services (February 2020)

Just before the publication of the White Paper, the EP adopted a resolution on automated decision-making processes.[245] In this resolution, the EP urges the EC to bring forward proposals to adapt the EU's safety rules both specific and general. It also further stresses the need for a risk-based approach and for a revision of the Product Liability Directive to ensure a functioning internal market ensuring clarity for private sector, trust, and protection for consumers.

Draft report with recommendations on a civil liability regime for AI (April 2020)

Two months later, the EP published a draft report with recommendations to the Commission on the adoption of a Civil liability regime for artificial intelligence.[246] This draft report goes further than the previously analysed documents under this section, as it formulates a genuine draft legislative text.[247] The report translates the past and current recommendations in concrete legal provisions.

Firstly, the report suggests adopting a principle-based, future proof and horizontal legal framework. For this purpose, the choice of a Regulation to ensure a proper harmonisation and common rules on AI systems in which the question of liability will be a key aspect. [248] The report also further confirms that there is no need for a complete revision of the well-functioning liability regimes, however, it also acknowledges that the characteristics of AI systems deserves specific rules including opacity. The report also confirms that in most cases the fault-based tort law of Member State offers a sufficient level of protection. The document further defines key notions

[245] European Parliament, 'Resolution on automated decision-making processes: ensuring consumer protection and free movement of goods and services', 2019/2915, 12.02.2020 https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.pdf.
[246] European Parliament, Draft Report with recommendation to the Commissions on a Civil Liability regime for artificial intelligence, 2020/2014, 27.04.2020, https://www.europarl.europa.eu/doceo/document/JURI-PR-650556_EN.pdf.
[247] ibid. p. 10-24.
[248] ibid. p. 5-6.

such as deployers[249] and indicates that in case of multiple deployers, all of them must be jointly and severally liable.[250]

Following what was advanced in the White Paper and the risk-based approach, the report here further builds upon this concept and recognizes that an AI-system that entails a high risk[251] potentially endangers the general public to a much higher degree and therefore should be subject to specific rules and advances setting up a strict liability regime.[252] The report also recommends listing all high-risk AI systems in an Annex and further clarifies that the proposed regulation should only cover harm to the important legally protected rights such as life, health, physical integrity and property, and should set out the amounts and extent of compensation as well as the limitation period.[253]

For non-high-risk systems, they should remain subject to fault-based liability.[254]

Researchers observed some shortcomings of the report such as the fact that the one-size-fit-all approach advanced would not consider existing (supra)national sectoral liability regimes and that the report refers to national law for the interpretation of legal concepts relating to liability which is undermining the harmonisation effect pursued.[255] The relationship between the Product Liability Directive is also not tackled, the annex listing the high risks sectors and systems does not list healthcare, and some notions are unclear such as 'deployer', which is broad but vague. [256]

Study on Liability for Artificial Intelligence (July 2020)

The Study realized by Andrea Bertolini provides interesting insights, sometimes departing from other studies on the same topic. Firstly, he underlines how difficult it is to define and classify AI. Mr. Bertolini regrets that current recommendations for AI and liability are technologically neutral as this approach pursues a one-size-fits-all dynamic which is not fitting the dynamic and

[249] Deployer or the person who decides on the use of the AI-system, who exercises control over the risk and who benefits from its operation.
[250] European Parliament, Draft Report , p.7.
[251] High risk AI system is defined in the report as when its autonomous operation involves a significant potential to cause harm to one or more persons, in a manner that is random and impossible to predict in advance; considers that the significance of the potential depends on the interplay between the severity of possible harm, the likelihood that the risk materializes and the manner in which the AI-system is being used.
[252] European Parliament, Draft Report , p.7.
[253] ibid. , p.7.
[254] Ibid.
[255] De Bruyne J., Dheu O., An EU Perspective on Liability and Artificial Intelligence, 14 May 2020 https://ai-laws.org/en/2020/05/an-eu-perspective-on-liability-and-artificial-intelligence/.
[256] ibid.

fast evolving AI systems.[257] Instead, he suggests adopting a harmonized and uniform European approach to AI and liability, revising the product safety liability in a more technology neutral approach and adopting more specific ad-hoc regimes to tackle all the challenges that diverse technologies are creating following a risks management approach. [258] This approach focusing on specific AI applications would be based on a strict no-fault liability, with a different liable actor depending on the application at stake but with a right of recourse against the other potential liable partners. [259]

The study suggests reforming the Product Liability Directive (PLD) to improve its implementation without clarifying further what should be revisited. But he points that the AI systems will challenge the PLD system at the detriment of the victim, which will struggle to obtain compensation. According to the study, the revised PLD should have a residual character besides ad hoc liability regimes.[260] Ad hoc regimes which would be based on a risk management approach "where the party best positioned to control or mitigate the risks would legally responsible".[261] The study also warns against a low and high risks regime as determining clearly the different liabilities is an almost impossible task given the lack of data on damages, the relevance of the criteria and risks of denied justice for low categorised AI systems.[262] The need to avoid under compensation of victims, having a single-entry point for litigation and a clear identification of the responsible party were underlined and show how the victim's compensation concerns are the key parameters for a successful revision of liability rules. Researchers pointed that the study does not clarify how this approach would clearly interplay with the PLD.

Resolution on a Civil Liability Regime for Artificial Intelligence (October 2020)

As already mentioned, the third resolution adopted by the EP in October 2020 was the Resolution 2020/2014(INL) on a Civil Liability Regime for Artificial Intelligence.[263] The resolution calls for a revision of the Product Liability Directive and legal certainty about the liability chain for AI systems. The EP considered that operator liability rules should apply to all types of AI system operations, regardless of the location of the operation and whether it is of a physical or virtual nature.

Compulsory insurance and strict liability for operators of a high-risk AI-system causing any harm or damage that was triggered by a physical or virtual activity, device or process driven by that

[257] Andrea B, Study on artificial intelligence and civil liability, (2020), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf, p. 15-31.
[258] ibid. p. 98.
[259] ibid. p. 91-123.
[260] Ibid. p. 124.
[261] ibid.
[262] Ibid, p. 63-81.
[263] European Parliament, Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)).

AI-system were put forward without the possibility to exonerate for due diligence action. For legal certainty, an exhaustive list of all high-risk AI systems should be set out in an annex to the forthcoming regulation and reviewed every 6 months to stay up to date with the rapid technological progress in this field. Regarding compensation, the EP suggests a maximum amount of EUR two million in case of death, harm to health or physical integrity (limitation period 30 years) and one million in case of significant immaterial harm that results in a verifiable economic loss or of damage caused to property (limitation period 10 years). The other non-high-risk AI systems should be governed by a fault-based liability with due diligence exoneration.

Conclusion

As demonstrated by the initiatives hereabove, liability and safety are a recurring theme, as evidenced by the review of some of the recent policy initiatives at the EU level in the field. Indeed, the essential characteristics of AI systems such as opacity, complex value chain and complex systems, autonomy, connectivity, data dependency can make extremely hard for victims to obtain compensation with traditional rules not reflecting these specificities in legal modalities. This is however tremendously important for the well-functioning of the internal market. Several options were put forward through reports and studies examined, but they all agreed that the current legal framework needs revisions.

The AI Act proposal (see Section 4.1.3) takes over and builds upon the work previously done at the EU level in order to formulate concretely some suggestions and move forward towards more trust for consumers but also from the private sector to operate in the EU internal market. Before that, supplementary policy initiatives will be analysed.

3.2.5 Other policy initiatives

As mentioned above, in October 2020, the EP adopted a number of resolutions related to AI, including on ethics, liability and copyright. In 2021, those were followed by resolutions on AI in criminal matters and in education, culture and the audio-visual sector.

3.2.5.1 AI and criminal law

In the motion for a resolution on AI in criminal law and its use by the police and judicial authorities in criminal matters (2020/2016(INI)) of 8 of June 2020, the EP recognizes that whereas AI applications offer great opportunities in the field of law enforcement, they also entail a number of potential risks, such as opaque decision-making, different types of discrimination, and risks to the protection of privacy and personal data, the protection of freedom of expression and information, and the presumption of innocence. The proposal for a resolution underlines that in judicial and law enforcement contexts, the final decision always needs to be taken by a human, who can be held accountable for the decisions made, and include the possibility of a recourse for a remedy. It also calls for algorithmic explainability and transparency of AI decisions. Finally, it calls for a moratorium on the deployment of facial recognition systems for law

enforcement, until the technical standards can be considered fully complaint with fundamental rights compliant.

3.2.5.2 Use of AI in education, culture and the audio-visual sector

On 19th May 2021, the EP adopted a resolution on AI in education, culture and the audiovisual sector. The resolution recognizes the following uses of AI systems in the media sector (Figure 13):

1. AI can help to promote linguistic diversity and support culture in the Union. It can contribute to the wider dissemination of European audiovisual works, in particular through automatic subtitling and dubbing of audiovisual content in other languages;
2. AI drives innovation in newsrooms by automating a variety of mundane tasks, interpreting data and even generating news such as weather forecasts and sports results;
3. AI creates new tools, software and AI-assisted production for easier content production;
4. AI provides tools to enable the broader public to create content;
5. AI technologies contribute to the creation, planning, management, production, distribution, localisation and consumption of audiovisual media products;
6. AI can be used to detect manipulated content such as deepfakes and combat such malicious activity through real-time fact checking, flagging, filtering out or labelling the content. AI can help save costs and drive innovation for the media sector in the direction towards a better allocation of the human forces and competence improving the quality and variety of content.

Figure 13: AI systems applications in the media landscape

The text acknowledges several issues and challenges that the development of AI encounters and wishes to establish clear guidelines for any AI progress for these sectors. The following analysis will focus primarily on media considering the project's focus.

First, the EP underlined the considerable influence of education, cultural programmes and audio-visual content in shaping people's beliefs and values. Therefore, any development, deployment and use of AI and related technologies in these sensitive sectors must fully respect the fundamental rights, freedoms and values enshrined in the EU treaties.

Second, the resolution appreciates the placing of AI and related technologies high on the agenda as they underlined the omni presence of AI applications in the audio-visual sector, in particular on audio-visual content platforms. It underlines how AI has already entered the creative value chain at the level of creation production, dissemination and consumption and is therefore having an immense impact on the culture and creative sectors and industries, including music, the film industry, art and literature.

The MEPs also further underlined several challenges related to AI in these fields, such as (Figure 14):

- The problem of the datasets composition which can lead to discrimination or replicate existing ones.
- The fact that 1 on 10 women or girls has been a victim of cyber harassment.
- For the audio-visual sector, the problem of data access from the global platforms and major players to ensure a level playing field.
- The need for adequate equipment and infrastructure.
- Not enough IT expertise, digital education, media training, digital skills.
- Risk of non-inclusivity. The MEPs underlined the importance to ensure inclusivity and that the whole of society is equally and fairly represented when developing, deploying, and using AI technologies
- Data protection issues, risks of discriminatory output based on biased input, risks for media and opinion pluralism, risks for cultural and linguistic diversity when AI technologies are used for cultural and creative content.
- Need to improve the accessibility of cultural and creative content for people with disabilities.
- AI-generated fake content such as deepfakes leading to disinformation, misinformation and hate speech campaigns and the lack of legal framework on this issue.
- The potentially negative impact of personalised advertising, microtargeted and behavioural advertising, and the assessment of individuals, especially minors.
- Lack of clarity on intellectual property rights, especially for the conditions of use of copyright-protected content as data input (images, music, films, databases, etc.) and in the production of cultural and audiovisual outputs, whether created by humans with the assistance of AI or autonomously generated by AI technologies.

Figure 14: Challenges created or facilitated by AI systems applications in the media landscape

To mitigate these challenges, the EP calls the EC to present a general regulatory framework, which applies to all applications of AI, and to complement it with sector-specific rules, for example for audio-visual media services. MEPs also call the EC to introduce strict limitations on targeted advertising based on the collection of personal data, starting with a ban on cross-platform behavioural advertising.

Moreover, the EP calls the EC and Member States to fully incorporate the gender and the diversity dimension in the measures taken related to AI: datasets, training, education, developers' team, research. Finally, the EP calls the EC in close cooperation with Member States and the relevant stakeholders, to develop verification mechanisms or systems for publishers, authors, and creators to assist them in verifying what content they may use and to determine more easily what is protected under IPR legislation.

The EP calls for the establishment of a clear ethical framework for the use of AI technologies in media to prevent all forms of discrimination and ensure access to culturally and linguistically

diverse content at Union level, based on accountable, transparent, and inclusive algorithms, while respecting individuals' choices and preferences. The EP supports the risk-based approach to be taken in the upcoming regulation on AI (see Section 4.1.3). The EP asks the EC to add education at the high-risk AI systems list.

The EP also focused on transparency and stressed the need for media organisations to be informed about the main parameters of algorithm-based AI systems that determine ranking and search results on third-party platforms. This is also for users to be informed about the use of AI in decision-making services and empowered to set their privacy parameters via transparent and understandable measures. This last point has already partially materialised in the Digital Services Act (DSA, see Section 4.2) proposal released earlier in December 2020, but the scope remains limited to some AI applications.

Furthermore, the EP calls for more research on the impact of AI on European creative industries and IP, online streaming services, and the risks of AI assisting the spread of disinformation in the digital environment, as well as solutions on how AI could be used to help counter disinformation.

AUDIO-VISUAL SECTOR

Moreover, specific recommendations have been made when it comes to the audio-visual sector. MEPs request that the algorithms used by media service providers, video sharing platforms (VSPs) and music streaming services should ensure that personalised suggestions do not put forward the most popular works, for targeted advertising, commercial purposes or to maximise profit. The EP calls for a recommendation on algorithms and personalised marketing striving for explainability, transparency and non-discriminatory outputs in line with the recently adopted Platform to Business Regulation[264] and the New Deal for Consumers Omnibus Directive.[265] It also proposes recommendations to increase user control over algorithms used for content recommendation with an option to opt-out from recommendation and personalised services. MEPs express their opinion that algorithms should only be used as a flagging mechanism in content moderation, subject to human intervention. At this occasion they recall that there should be no general monitoring according to article 15 of the E-commerce Directive. Finally, the EP further calls for the development of indicators to assess cultural diversity and the promotion of European works on such services. Linked to what is enshrined to the DMA, the MEPs have also stressed the importance to take regulatory measures to ensure that media service providers have access to the data generated by the provision and dissemination of their content on other providers platforms.

[264] Regulation (EU) 2019/1150 of the European Parliament and of the Council of 20 June 2019 on promoting fairness and transparency for business users of online intermediation services (OJ L 186, 11.7.2019, p. 57).

[265] Directive (EU) 2019/2161 of the European Parliament and of the Council of 27 November 2019 amending Council Directive 93/13/EEC and Directives 98/6/EC, 2005/29/EC and 2011/83/EU of the European Parliament and of the Council as regards the better enforcement and modernisation of Union consumer protection rules, (OJ L 328, 18.12.2019, p. 7).

ONLINE DISINFORMATION: DEEPFAKES

The EP recalls that accuracy, independence, fairness, confidentiality, humanity, accountability and transparency, as driving forces behind the principles of freedom of expression and access to information in online and offline media, are decisive in the fight against disinformation and misinformation. It encourages EC to continue its work on disinformation and creating awareness about this problem and major issue and asks for more educational measures in this regard.

The EP asks for an appropriate legal framework to govern deepfakes creation, production, or distribution for malicious purposes, and to propose recommendations for, among other initiatives, action against any AI-powered threats to free and fair elections and democracy. In addition, it also calls the EC to impose an obligation for all deepfake material or any other realistically made synthetic material to state that the material is not original, and a strict limitation when used for electoral purposes. This resolution asks for further requirements than the one currently present in the AI Act proposal (see Section 4.1.3).

To fight filter bubbles and echo chambers, which are restricting diversity of opinion and undermining open debate in society, the EP urges that the transparency for the algorithms to process information must be ensured. Users must be empowered and given greater freedom to decide whether and what information they want to receive. When used to generated content such as automated news articles, the EP underlines that the quality of the datasets and editorial supervision must be ensured.

3.2.5.3 Technology AI Policy Initiatives

Finally, there are also a number of various EU initiatives in the field of cloud computing and quantum technologies. It is worth mentioning that starting in September 2020, the GAIA-X imitative aims to strengthen Europe's position in data-driven innovations by developing common requirements for a European data infrastructure based on European values. Such an open data infrastructure should enable a secure, federated system that meets the highest standards of digital sovereignty while promoting innovation. The GAIA-X Community consists of companies and organisations that actively participate in the development of GAIA-X and that uphold the European values of enhanced data privacy, transparency, security and respect for data rights. The community works on open-source software and builds a common infrastructure. To address the challenges of a trustworthy data-driven economy, GAIA-X builds on cloud solution providers, high performance computing (HPC) and edge systems. GAIA-X identifies the minimum technical requirements and services necessary to operate the federated GAIA-X Ecosystem. The development of these services will follow the principles of Security by Design and also include the concept of Privacy by Design in order to ensure highest security requirements and privacy protection.

The Quantum Technologies Flagship is a long-term research and innovation initiative that aims to develop quantum technologies. It specifically funds projects on quantum simulation,

quantum communication, quantum metrology and sensing, and quantum computing. The latter aims to leverage the new capabilities of quantum approaches to solve otherwise insoluble problems, to process vast amounts of data faster to recognize patterns and train AI systems. Indeed, computational power is a main challenge of modern AI approaches, that are highly greedy in computations and hence in electrical power, entailing crucial challenges in terms of access to adequately powerful computational resources, and in terms of environmental impact. Quantum approaches open up new directions to tackle these challenges.

4 EU Regulatory initiatives in the field of AI

During the past year, the EC building on the various policy initiatives presented in Section 3, proposed a comprehensive package of regulatory measures that address problems posed by the development and use of AI and digital platforms. These include the AI Package, the Digital Services Act and the Digital Markets Act, as well as the Data Governance Act and the forthcoming Data Act.

4.1 AI Package

On 21 April 2021, the Commission published its AI package proposing new rules and actions aiming to turn Europe into the global hub for trustworthy AI. This includes:

- Communication on Fostering a European Approach to Artificial Intelligence (Section 4.1.1);
- Coordinated Plan on Artificial Intelligence 2021 Review (see Section 4.1.2);
- Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (see Section 4.1.3).

4.1.1 Communication on Fostering a European Approach to Artificial Intelligence

In the Communication, the Commission notes that the AI package represents a key milestone on the way to a European approach to AI. It is the outcome of three years of intense policymaking on AI at European level. As set out in the White Paper on AI, and largely confirmed by the public consultation that followed, the use of AI creates a number of specific high risks for which existing legislation is insufficient. To be future-proof and innovation friendly, the proposed legal framework is designed to intervene only where this is strictly needed and proposes a proportionate and risk-based European regulatory approach. The Commission also notes that the proposed regulatory framework on AI will work in tandem with applicable product safety legislation and in particular the revision of the Machinery Directive, the recently proposed Digital Services Act and Digital Markets Act (see Section 4.2) as well as the European Democracy Action Plan. Finally, the proposed framework will be complemented by legislation to adapt the EU liability framework, such as revising the Product Liability Directive, in order to address liability issues related to new technologies, including AI, and by a revision of the General Product Safety Directive.

The detailed assessment of the Coordinated Plan on Artificial Intelligence 2021 Review and the Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act, AI Act) can be found below.

4.1.2 Coordinated Plan on Artificial Intelligence 2021 Review

The 2021 review of the Coordinated Plan[266], published on 21 April 2021, puts forward a concrete set of joint actions for the EC and Member States on how to create EU global leadership on trustworthy AI. It puts forward a vision to accelerate investments in AI, spur the implementation of national AI strategies and align AI policy to remove fragmentation and address global challenges. The review of the Coordinated Plan puts forward four key sets of proposals for the EU and the Member States:

Set enabling conditions for AI's development and uptake in the EU

The updated Plan identifies the following enabling conditions which are necessary in order to support the development and take-up of AI. First, an appropriate governance and coordination framework. The EC notes that Member States made substantial efforts to develop national strategies on AI. This helped the EC to identify the priority sectors for joint actions and provided a solid mapping of countries' priorities. Now, the EC commits to further facilitate the update of and synergies between national actions, e.g. through AI Watch[267] and encourages Member States to review and update national AI strategies, develop and promote instruments to allow regular monitoring, coordination and exchange between stakeholders, reinforce support and investment in AI, and share and develop actions on national/regional level, which were successful in other Member States (e.g. on a virtual warehouse of data). The EC also recalls that it has established three horizontal expert groups: (i) High-Level Expert Group on Artificial Intelligence; (ii) High-Level Expert Group on the Impact of the Digital Transformation on EU Labour Markets; (iii) Expert Group on Liability and New Technologies (see Section 2.4.1. of this deliverable). In addition to these horizontal groups, numerous sectoral expert groups focused on specific policy areas (e.g. autonomous vehicles, aviation, mobility and transport, home affairs and emerging security risks). Among the various initiatives listed in the Coordinated Plan, the AI Alliance[268] is worth mentioning. The AI Alliance is an online forum, set up by the EC to engage more broadly with stakeholders on AI-related topics. It provides a platform currently gathering around 4,000 stakeholders from all around the world to exchange information and discuss the technological and societal implications of AI. In the Coordinated Plan the EC commits to continue collecting data on AI developments, organising annual AI Assemblies, and will by 2022 propose how to reinforce monitoring of the developments and impact of AI technologies in the EU. It will also regularly monitor the implementation of the Coordinated Plan. In order to facilitate governance mechanisms for cooperation, the Member States' Group on AI and Digitising

[266] 'European Commission, Coordinated Plan on Artificial Intelligence 2021 Review COM(2021) 205 Final'.
[267] In 2018 the Commission launched AI Watch in order to monitor developments relating to AI technologies. AI Watch (run by the Commission's Joint Research Centre (JRC)) has worked in coordination with the Member States.
[268] For an overview of the AI Alliance, see European Commission, The European AI Alliance (information webpage, 2020).

European Industry, facilitated by the EC, will continue to steer discussions between Member States and the EC.

Second, not surprisingly, the EC notes that the development of AI technologies often requires large, high quality, secure and robust datasets.[269] In the EC's words, "the availability of high-quality data, among other things, in respect of diversity, non-discrimination, and the possibility to use, combine and re-use data from various sources in a GDPR compliant way are essential prerequisites and a precondition for the development and deployment of certain AI systems."[270] In that respect, the EC has published various data-related initiatives, including the European strategy for data[271] adopted on 19 February 2020 and the new Data Governance Act[272] proposed on 25 November 2020 (see Section 4.3). To further support actions on data, the EC will: adopt a proposal for a Data Act, in order to stimulate the use of privately-held data by government (B2G), address issues related to data access and use in business-to-business settings, in particular non-personal data resulting from objects connected to the internet of things (Q3 2021); at the time of writing, the impact assessment is available.[273] The EC is also expected to propose an implementing act on making public sector high-value data sets in a machine-readable format freely available for reuse (Q2 2021).[274]

Other EC initiatives include launching of a European Alliance for industrial data, edge and cloud, investment in European data spaces and the European cloud federation. The EC intends to build European data spaces for various sectors, including industrial manufacturing, the green deal, mobility, health, finance, energy, agriculture, public administration, and skills. Interestingly, although not mentioned either by the Data Strategy or by the updated Coordinated Plan, the EC also proposes the creation of a European "media data space".[275]

Third, a computation infrastructure. This infrastructure is necessary for storing, analysing and processing the increasingly large volumes of data, which in turn, requires new developments and approaches to increase computing capabilities. The Commission has already taken a number of measures to support the development of the technological systems and the next-generation of data processing infrastructures. In particular, to support the development of High Performance Computing (HPC) capabilities in Europe, the EuroHPC Joint Undertaking coordinates efforts and pools resources among 32 participating countries to develop and deploy

[269] European Commission, Coordinated Plan on Artificial Intelligence 2021 Review COM(2021) 205 Final (n 269).
[270] ibid.
[271] European Commission, A European strategy for data, (COM(2020) 66 final).
[272] European Commission, Proposal for a Regulation on European data governance (COM(2020) 767 final).
[273] European Commission, Data Act Inception impact assessment - Ares(2021)3527151.
[274] Based on Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (OJ L 172, 26.6.2019, p. 56).
[275] See also Section 4 of this deliverable.

a world-class supercomputing infrastructure that will be easily and securely accessible from anywhere in Europe. The EC with the support of the Member States will launch an Industrial Alliance on Microelectronics[276] and invest in research and innovation for the computing needs of low-power edge AI). The EC also encourages Member States to continue the development of national integrated large-scale data management and HPC infrastructure and invest in strengthening Europe's position in processors and semiconductor technologies for AI.

Make the EU the place where excellence thrives from the lab to the market

Key proposals for excellence include: First, to collaborate with stakeholders through, e.g. the European Partnership on AI, Data and Robotics and expert groups. The EC will continue to support European partnerships in the context of Horizon Europe and enhance the strategic approach to research and innovation (R&I) in AI technologies. The EC will, in 2021 support a number of initiatives, including a co-programmed European partnership on AI, Data and Robotics, the co-programmed European Partnership on Photonics, the co-programmed European Partnership 'Made in Europe' as well as support and facilitate synergies between European partnerships.

Second, to build and mobilise research capacities. Through Horizon 2020, the EC invested EUR 50 million over 4 years to create a research community of closely networked AI excellence centres. Moreover, starting in 2021, the EC will together with the Member States and the wider AI community, set up an AI lighthouse for Europe, as announced in the White Paper. The AI lighthouse will build on the existing and future Networks of AI excellence centres, with the aim to build an alliance of strong European research organisations that will share a common roadmap to support excellence in basic and applied research, to align national AI efforts, to foster innovation and investments, to attract and retain AI talent in Europe, and to create synergies and economies of scale. The EC also devoted to fund, under Horizon Europe, in 2021 and 2022, additional networks of AI excellence centres and advance the state of the art in various areas of AI research, including research towards the next level of intelligence and autonomy of AI based systems, transparency in AI, greener AI, AI for complex systems, advances in edge AI networks, unbiased AI systems, empowering humans with advanced AI support. The EC has also committed on the aim that AI-related projects that receive R&I funding under the Horizon Europe adhere, as appropriate, to the 'ethics by design' principle, including for trustworthy AI.

Third, to provide tools through an AI-on-demand platform and an environment for developers to test and experiment (TEFs), and for SMEs and public administrations to take up AI (EDIH). TEFs are technology infrastructures with specific expertise and experience in testing mature technology in a given sector, in real or close-to-real conditions; and EDIHs 'one-stop shops' that help all companies interested to use AI to become more competitive with regard to their business/production processes, products or services by using AI technologies. In order to help

[276] See 'Joint declaration on processors and semiconductor technologies', available at: <https://digital-strategy.ec.europa.eu/en/library/joint-declaration-processors-and-semiconductor-technologies>.

to bring innovation from the 'lab to the market' – to ensure the broad uptake and deployment of AI technologies, the EC together with Member States will: (i) co-fund Testing and Experimentation Facilities under the Digital Europe programme. In this context the first calls (in 2021-2022) will focus on the following identified sectors: manufacturing, health, agri-food, smart communities and edge AI. The estimated budget per sector will be around EUR 20-75 million; (ii) select, during 2021-2022, the network of up to 210 EDIHs covering all regions of Europe; (iii) consolidate in 2021 and onwards, the AI-on-demand platform as a central European toolbox of AI resources needed for industry and public sector.

Fourth, to fund and scale innovative ideas and solutions for AI. Key EC actions under this heading include: (i) strengthening the support and funding for the AI/Blockchain Investment pilot and support programme through the InvestEU programme 2021-2027, and (ii) fully implementing the European Innovation Council (EIC) under Horizon Europe. In addition, (iii) the Women TechEU initiative to be launched to support deep-tech start-ups founded and led by women; (iv) mobilizing AI start-ups in national hubs through Startup Europe and the Innovation Radar; (v) facilitating exchange of information, expertise and best practice between local, regional and national AI start-ups at European level, and (vi) the availability of open data and access to data for SMEs.

Ensure that AI works for people and is a force for good in society

Key proposals to ensure that AI works for people include:

(1) Nurture talent and improve the supply of skills necessary to enable a thriving AI eco-system. The 2018 Coordinated Plan identified the significant ICT skills gap as a key challenge to the development of AI in Europe. In that respect, the EC adopted in September 2020 a new Digital Education Action Plan for the period 2021-2027. As a part of the actions planned in the Education Plan, the EC commits to support traineeships in digital areas, develop ethical guidelines on AI and data usage in teaching and learning for educators as well as the support of related research and innovation activities through Horizon Europe (e.g. networks of AI excellence centres will explore options to retain talents and develop PhD programmes in AI, which could be integrated in non-ICT education). Under the Digital Europe programme, the call for the specialised programmes will be launched in Q1/2 2021 and the short-term training courses in Q1 2022. Member States are equally encouraged to refine and implement the skills dimension in their national AI strategies and take measures and exchange best practices on how to integrate AI into general education and how to increase inclusion and diversity.

(2) Develop a policy framework to ensure trust in AI systems. Trust in essential to facilitate the uptake of AI technologies. In that respect, the AI HLEG has identified key principles and requirements for Trustworthy AI (see Section 3.2.2.1. of this deliverable). Other actions taken by the EU include adopting the EU Cybersecurity Strategy for the Digital Decade, and the Intellectual Property action plan. In the Coordinated Plan, the EC commits to propose a legislative action on a horizontal framework for AI, to propose in 2021 and onwards revisions of

existing sectoral safety legislation, and to propose in 2022 EU measures adapting the liability framework to the challenges of new technologies, including AI. It also commits to organise stakeholders dialogues, further strengthen cooperation with EU agencies and other relevant EU bodies and other organisations.

(3) Promote the EU vision on sustainable and trustworthy AI in the world. The EU actively participates in global AI initiatives e.g. the EU is a founding member of the Global Partnership in AI (GPAI), it contributes to the OECD's work on AI, supports international standardization bodies such as the ISO and the IEEE and participates in numerous bilateral dialogues with third countries such as Japan or Canada. The EC commits to continue its participation and support to these international, multilateral and bilateral discussions on trustworthy AI and to foster the setting of global AI standards.

Build strategic leadership in high-impact sectors

In addition to the horizontal actions, the 2021 review of the Coordinated Plan also proposes seven sectoral action areas: (1) Bring AI into play for climate and environment; (2) Use the next generation of AI to improve health; (3) Maintain Europe's lead: Strategy for Robotics in the world of AI; (4) Make the public sector a trailblazer for using AI; (5) Apply AI to law enforcement, migration and asylum; (6) Make mobility safer and less polluting through AI; (7) Support AI for sustainable Agriculture. For details of each of these sectoral action areas, we refer to the Coordinated Plan.

Conclusion

The revised Coordinated Plan sums up that the objectives of the 2018 Coordinated Plan remain relevant and the overall direction set in the Coordinated Plan has proven to be the right one to contribute to Europe's ambition to play the world-leading role in developing ethical, trustworthy and human-centric AI. However, the EC notes that the next steps should focus on the implementation of the joint actions and the removal of fragmentation between funding programmes, initiatives and actions taken at EU and Member State level. The EC will, in collaboration with the Member States, closely monitor and follow up on the progress made in the implementation of the joint actions agreed in the Coordinated Plan.

4.1.3 Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)

The proposal comes in accordance with the political commitment made by President von der Leyen in her Political Guidelines.[277] Its primary objective is to ensure the proper functioning of the internal market by setting harmonised rules in particular on the development, placing on

[277] European Commission, A Union that strives for more, My agenda for Europe : political guidelines for the next European Commission 2019-2024, < https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf>, accessed 15 July 2021.

the Union market, and use of products and services making use of AI technologies or provided as stand-alone AI systems. As provided by the Explanatory Memorandum, the Regulation has the following specific objectives:

(i) ensure that AI systems placed on the Union market and used are safe and respect existing law on fundamental rights and Union values;
(ii) ensure legal certainty to facilitate investment and innovation in AI;
(iii) enhance governance and effective enforcement of existing law on fundamental rights and safety requirements applicable to AI systems;
(iv) facilitate the development of a single market for lawful, safe and trustworthy AI applications and prevent market fragmentation.

Risk-based approach

The regulation follows a risk-based approach, differentiating between uses of AI that create (i) unacceptable risks (Title II); (ii) high risks (Title III); (iii) limited risks (Title IV); and (iv) minimal risks (Title IX) (see Figure 15 below).

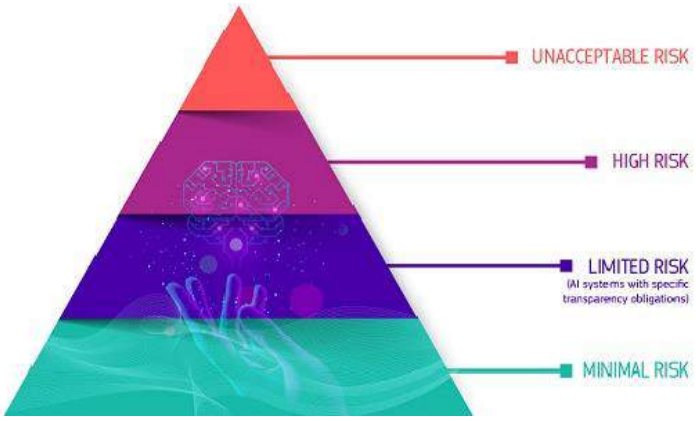

Figure 15: A risk-based approach to AI regulation[278]

Analysis of selected provisions

In what follows, we will briefly comment on the selected provisions with a focus on those potentially applicable for media applications. Providing a detailed analysis of the proposal for the Regulation is beyond the scope of this document and will be addressed in Deliverable D2.4 on Pilot Policy recommendations in the field of AI and media.

[278] European Commission, New rules for Artificial Intelligence – Facts page, <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en> accessed 10 June 2021.

SCOPE AND DEFINITIONS (TITLE I)

First, Title I defines the subject matter of the regulation. It lays down harmonized rules for the placing on the market, the putting into service and the use of artificial intelligence systems ('AI systems') in the Union; prohibitions of certain AI practices; specific requirements for high-risk AI systems and obligations for operators of such systems; harmonised transparency rules for certain AI systems; and rules on market monitoring and surveillance.[279] The definitions of 'placing on the market' (the first making available of an AI system on the Union market)[280] and 'putting into service' (the supply of an AI system for first use directly to the user or for own use on the Union market for its intended purpose)[281] make it unclear whether, and to what extent, the proposed AI regulation is applicable to scientific research activities (such as within AI4Media). More on this aspect can be found in Section 5 of this deliverable (The potential impact of the anticipated EU regulatory initiatives in the field of AI for AI4Media project).

Second, as provided by Recital 11, in light of their digital nature, certain AI systems should fall within the scope of the regulation even when they are neither placed on the market, nor put into service, nor used in the Union. To sum up, the proposed regulation applies to: providers placing on the market or putting into service AI systems in the Union, irrespective of whether those providers are established within the Union or in a third country; users of AI systems located within the Union; and providers and users of AI systems that are located in a third country, where the output produced by the system is used in the Union.[282]

Third, Title I also sets out the definitions used throughout the regulation. Importantly, both the definition of a 'provider'[283] and a 'user'[284] are defined broadly and seem to encompass a variety of entities such as universities, research centers, as well as media agencies. 'Artificial intelligence system' is defined as software that is developed with one or more of the techniques and approaches listed in Annex I of the Regulation and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. The definition resembles to a great extent the

[279] Art. 1 of the AI Act.
[280] Art. 3(1)(9) of the AI Act.
[281] Art. 3(1)(11) of the AI Act.
[282] Art. 2(1) of the AI Act.
[283] 'provider' means a natural or legal person, public authority, agency or other body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark, whether for payment or free of charge.
[284] 'user' means any natural or legal person, public authority, agency or other body using an AI system under its authority, except where the AI system is used in the course of a personal non-professional activity.

OECD's Recommendation of the Council on Artificial Intelligence[285] definition. There are, however, two caveats worth exploring. First, AI system is defined as a 'software', and not as a 'machine-based system' as defined by the OECD. 'Hardware', on the other hand, is not covered by the AI Regulation. However, on the 21st of April 2021, the EC presented its proposal for a new Regulation on machinery products.[286] The new Machinery Regulation intends to ensure the safe integration of the AI system into the overall machinery. Both AI systems which are used on a stand-alone basis and as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serves the functionality of the product without being integrated therein (non-embedded), are covered by the Regulation. Second, the definition is complemented by Annex I, which contains a detailed list of approaches and techniques.[287]

Prohibited artificial intelligence practices (Title II)

Title II establishes a list of prohibited AI practices. Importantly, the draft Regulation does not prohibit AI systems *as such*, but only *practices* or, in other words, certain applications of the AI systems. Article 5 comprises AI systems whose use is considered unacceptable as contravening Union values, such as human dignity, freedom, equality, democracy, the rule of law, and union fundamental rights prescribed in the EU Charter of Fundamental Rights. The prohibitions enumerate the following as prohibited practices:

Subliminal Practices:

"(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness in order to materially distort a person's behavior in a manner that causes or is likely to cause that person or another person physical or psychological harm."[288] In this provision, the notion of 'subliminal techniques' is not defined, which makes the scope of application of this provision far from clear. One may wonder whether and to which extent the online social media practices such as dark patterns fall within the scope of this provision. Additionally, the provision requires a person's behavior to be "materially

[285] AI system: An AI system is a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy.
https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

[286] European Commission, COM(2021) 202 - Proposal for a Regulation of the European Parliament and of the Council on machinery products

[287] The list includes: a) Machine learning approaches, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; (b) Logic- and knowledge-based approaches, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; (c) Statistical approaches, Bayesian estimation, search and optimization methods.

[288] Art. 5(1)(a) of the AI Act.

distorted". It is ambiguous what this concept – which seems to be borrower from the consumer protection legislation - would mean in the context of AI systems. As mentioned below, the "harm" criterion also raises some criticism.

EXPLOITING VULNERABILITIES OF SPECIFIC GROUPS:

"(b) the placing on the market, putting into service or use of an AI system that exploits the vulnerabilities of a specific group of persons due to their age, physical or mental disability, in order to materially distort the behavior of a person pertaining to that group in a manner that causes or is likely to cause that person or another person physical or psychological harm."[289]

Concerning this provision, first, it is important to note, that the applicability of it is limited to 'the vulnerabilities of a specific *group* of persons' (own emphasis). One could wonder whether a more individual-oriented approach would be needed. Moreover, the provision refers only to the vulnerabilities of a group of persons due to their 'age, physical or mental disability', leaving other characteristics (i.e. gender) or socio-economic vulnerable people who can be adversely affected by AI systems, aside. Additionally, the Explanatory Memorandum clearly distinguishes between manipulative or exploitative practices targeting or affecting certain specific groups and adults that are not considered within a specific group category. Therefore, practices targeting or affecting the latter shall still be covered by the existing data protection, consumer protection, and digital service legislation that guarantee that the natural persons are properly informed and have free choice not to be subject to profiling or other practices that might affect their behavior. Second, the scope of this provision is limited to those exploitation of certain vulnerabilities, that 'causes or is likely to cause that person or another person *physical or psychological harm'* (own emphasis). The harm requirement has already sparked some criticism.[290] In particular, manipulative AI systems appear permitted insofar as they are unlikely to cause an individual (not a collective) 'harm'.[291] The question raises whether the harm test should not be replaced by other alternative criteria, such as, for example a 'reasonable person' requirement.

Finally, Recital 16 provides for an exception by stating the intention to distort may not be presumed if the distortion of human behavior results from factors external to the AI systems which are outside of the control of the provider or the user. Though this exception sheds a tiny light on the provision's policy objective, the definition of "distortion" or what it entails has not enjoyed much clarity within the proposed regulation.

[289] Art. 5(1)(b) of the AI Act.
[290] Veale M. and Borgesius F.Z., 'Demystifying the Draft EU Artificial Intelligence Act' (SocArXiv 2021) preprint <https://osf.io/38p5f> accessed 20 July 2021.
[291] ibid.

SOCIAL SCORING BY PUBLIC AUTHORITIES:

The proposal also prohibits AI-based so-called 'social scoring' for general purposes used by or on behalf of public authorities (and not by private entities). The EDPB has already noted that any kind of social scoring should be prohibited, not only done 'over a certain period of time' or 'by public authorities or on their behalf' (so also pertaining to e.g. social media, cloud service providers).[292]

Interestingly, private scoring is not enumerated as a prohibited practice under this provision. However, the AI HLEG, whose guidelines are covered in details in Section 3.2.2.1 of this deliverable, found that any form of citizen scoring – whether conducted by public authorities or private actors – can endanger principles of human autonomy and non-discrimination. Thus, the AI HLEG suggests that if their practices impact human rights, private social scoring should also be banned.[293] Consequently, one can wonder whether the proposed regulation will include an amendment banning private social scoring, in context that impacts human rights, with its final version.

REAL-TIME BIOMETRIC IDENTIFICATION SYSTEMS FOR THE PURPOSES OF LAW ENFORCEMENT

Finally, the use of 'real time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement is also prohibited unless certain limited exceptions apply. Usage of real-time biometric systems in law enforcement is not within the scope of AI4Media's research. Hence, this deliverable only refers to this prohibited practice to provide brief compact information regarding the proposed regulation.

High-risk AI systems (Title III)

The Title III governs AI systems which pose 'high-risk' to 'health, safety and fundamental rights'[294] in certain defined applications, sectors and products. Title III applies to two main subcategories of AI systems: (i) the AI systems intended to be used as a safety component of a product, or is itself a product, already covered by the Union harmonisation legislation listed in Annex II (such as toys, machinery, lifts, or medical devices); (ii) stand-alone AI systems with mainly fundamental rights implications that are explicitly listed in Annex III. They include AI systems in eight fixed areas:

(i) biometric identification and categorisation (both 'remote', as in Title II above, and applied 'post' the event);

(ii) management and operation of critical infrastructure;

[292] EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 18 June 2021, https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf .
[293] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).
[294] Rec. 43, Art. 7(2) of the AI Act.

(iii) educational and vocational training;
(iv) employment, worker management and access to self-employment;
(v) access to and enjoyment of essential services and benefits;
(vi) law enforcement;
(vii) migration, asylum and border management;
(viii) administration of justice and democracy.

The EC can expand the list of high-risk AI systems used within certain pre-defined areas, but cannot add completely new areas.

The Act contains an extensive list of essential requirements (Chapter 2), which connects to obligations of regulated actors (Chapter 3). The vast majority of all obligations fall on the provider: person or body that develops an AI system or that has an AI system developed with a view to placing it on the market or putting it into service under its own name or trademark. Chapter 2 sets out the legal requirements for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security. The proposed requirements derive from the Ethics Guidelines of the AI HLEG (see Section 3.2.2.1 above). The detailed assessment of the requirements for a high-risk AI system is beyond the scope of this deliverable. However, the following issues are worth noting. First, datasets to train high-risk AI systems must meet data quality criteria, including in relation to relevance, representativeness, accuracy, completeness. Moreover, the proposed Act puts a *prima facie* high standard of datasets being 'free of errors and complete'. Providers must also ensure human oversight, incorporating 'human-machine interface tools' to ensure systems 'can be effectively overseen by natural persons'. Finally, it is important to note that the requirements for high-risk AI systems do not exist in legal vacuum. Some, such as privacy and explainability or non-discrimination, are dealt within already existing and applicable EU legislation, such as respectively the GDPR and the European Charter on Fundamental rights.

Chapter 3 places a set of horizontal obligations on providers of high-risk AI systems. Some obligations are also placed on users and other participants across the AI value chain (e.g., importers, distributors, authorized representatives). These include having a quality management system in place, drawing-up the technical documentation of the high-risk AI system and keeping the logs automatically generated by their high-risk AI systems. Moreover, providers must ensure that the high-risk AI system undergoes the relevant conformity assessment procedure, prior to its placing on the market or putting into service. Also, according to Article 51, before placing on the market or putting into service a high-risk AI system, the provider or, where applicable, the authorised representative shall register that system in the EU database. In order to indicate the conformity with the Regulation, high-risk AI systems should also obtain the CE marking. Chapter 4 sets the framework for notified bodies to be involved as

independent third parties in conformity assessment procedures, while Chapter 5 explains in detail the conformity assessment procedures to be followed for each type of high-risk AI system.

Transparency obligations for certain AI systems (Title IV)

Article 52 of the proposal sets transparency obligations which will apply for systems that (i) interact with humans, (ii) are used to detect emotions or determine association with (social) categories based on biometric data, or (iii) generate or manipulate content ('deep fakes').

Firstly, it is not clear why paragraph 1 is targeting 'providers' of the AI systems whereas paragraphs 2 and 3, 'users'. According to Article 52(1), providers of AI systems intended to interact with natural persons ('bots') must design and develop their systems in such a way that individuals are informed they are interacting with a bot, unless it would be obvious from the circumstances and the context of use, or if the bot use is authorised by law to detect, prevent, investigate and prosecute criminal offences. Two issues come to the fore. First, the draft AI Act does not provide any example as to when the interaction with a bot is 'contextually obvious', leaving some room for interpretation. Second, as mentioned above, according to the AI HLEG transparency requirement, humans should always be informed that they are interacting with an AI system. However, the AI HLEG Guidelines also mention that 'they should also have the option to have a human interaction instead.'[295] The draft AI Act, however, does not build on this statement and does create any "opt-out" right for individuals. The scope of this provision is also not so clear for the media sector, i.e. does the 'AI systems intended to interact with natural persons' encompass recommender systems or robot journalism?

Secondly, Art. 52(2) provides that "users of an emotion recognition system or a biometric categorisation system shall inform of the operation of the system the natural persons exposed thereto. This obligation shall not apply to AI systems used for biometric categorisation, which are permitted by law to detect, prevent and investigate criminal offences." Users are according to Art. 3(4) of the proposal "any natural or legal person, public authority, agency or other body using an AI system under its authority". Art. 52(2) provides an exception as where the AI system is used during a personal non-professional activity, the transparency requirement will not be binding. Recital 70 adds that the information and notifications should be provided in accessible formats for persons with disabilities.

Thirdly, users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake') shall disclose that the content has been artificially generated or manipulated. Recital 70 clarifies that the disclosure should be done by labelling the AI output accordingly and disclosing its artificial origin. This obligation,

[295] European Commission, The High-Level Expert Group on AI, Ethics Guidelines for Trustworthy AI, 2019 (n 31).

however, does not apply when deepfakes are used during a personal non-professional activity or where the use is authorised by law to detect, prevent, investigate and prosecute criminal offences or it is necessary for the exercise of the right to freedom of expression and the right to freedom of the arts and sciences guaranteed in the Charter of Fundamental Rights of the EU, and subject to appropriate safeguards for the rights and freedoms of third parties. Nor the recital nor the impact assessment report clarifies or illustrates what could be the scope of this paragraph. Freedom of the art implies that the arts and scientific research shall be free of constraint and the academic freedom shall be respected. It is not clear whether this implies that research on deepfake is exempt from these transparency requirements. When it comes to freedom of expression, this freedom enjoys a broad scope of protection even if it is not an absolute right. Some can wonder then what is the scope and impact of this provision as many deepfakes stories could be made and disseminated under the cover of freedom of expression. It can be argued that Art. 52(3) should also contain an obligation to make the deepfake identification information undeletable in case of transfer or further modification of the material in order not to lose track of the deepfake's original information.

The transparency requirements could include more precisions on what should be communicated (the type of information), when (at which stage this should be revealed) and how. Providing clearer guidance will help to gain a better understanding of the obligations for the users and providers and ensure a better and harmonized implementation of the provisions.

Measures in support of innovation (Title V)

In this Title, the EC aims to support innovation, research and smaller structures such as SMEs and start-ups. Article 53 and 54 provide information about the creation of sandboxes for AI established at the Member states level. The sandboxes are controlled frameworks in which innovative technologies can be tested according to defined parameters, conditions and time limits decided with the authorities. The provisions encourage the creation of such protected framework to experiment innovation and indicate, which measures in terms of governance, supervision and liability should be implemented for their materialisation. However, at this stage some obligations are not clarified in the provisions, which leaves the choice to Member States. Some fear that this might lead to a fragmented implementation of the regulatory sandboxes across the EU Member States and a race to the bottom with huge demand for the Member States with less restrictive rules.

Art. 55 focuses specifically on measures supporting SMEs by imposing obligations on Member States to provide them priority access to AI regulatory sandboxes, develop dedicated communication channels and awareness activities and take into account their specific needs and interest when setting conformity assessment fees.

Governance and implementation (Titles VI, VII and VIII)

Titles VI, VII, and VIII in conjunction set out a governance and implementation structure, in the form of establishing authorities for conformity assessment, oversight, and market surveillance. First, Art. 56 proposes to establish a 'European Artificial Intelligence Board' composed of national supervisory authorities. This board will aid in facilitation and effective harmonization of the implementation of the AI Act. Second, Art. 59 compels Member States to designate national authorities to supervise the implementation and application of the Act. Third, Art. 60 introduces a new EU-wide database for stand-alone high-risk AI systems to facilitate the monitoring work of the Commission and national authorities. Finally, Articles 61-68 set out a post-ante mechanism and monitoring and reporting obligations for providers of AI systems to ensure post-market surveillance, share of information on incidents and malfunctioning, and investigate compliance with the obligations and requirements for all high-risk AI systems already placed on the market. Additionally, Member States, with their existing sectorial authorities, will be the ones also having the authority to monitor and enforce the provisions of the regulation.

Codes of conduct (Title IX)

Article 69, the unique provision of this Title provides that both the EU and the Member States shall encourage and facilitate the creation of Codes of conduct intended to foster the voluntary application of the requirement for high-risks systems to AI systems falling outside this category, hence to all AI systems. The codes of conduct can also focus on specific requirements such as environmental sustainability or diversity in developers' team. Codes can also be created and implemented by providers and business association themselves.

Final provisions (Titles X, XI and XII)

Final provisions of the Act touch upon administrative issues such as confidentiality, penalties, and some obligation for the EC during the implementation phase. Art. 70 requires all the parties involved in the implementation of the Act to respect confidentiality of information and data obtained during the process and thereafter. Furthermore, Art. 71 introduces a penalty regime to ensure that all necessary measures are taken for the proper and effective implementation of the Act. The penalties, therefore, should be effective, proportionate, and dissuasive, also taking into consideration the interests of small-scale providers and start-ups. Thus, the Act foresees a three-tier penalty system, imposing higher administrative fines for non-compliance with prohibited practices and data governance obligations. Other compliance (second tier) and the supply of incorrect, incomplete, or misleading information (third tier) is set out be fined relatively lower than the first tier.

Lastly, the proposal lays down rules empowering the EC to adopt, implements acts or delegated acts concerning the update or complementation Annexes I to VII, if necessary, to ensure harmonised application of the regulation. This power comes with an obligation to regularly assess the need for an update of Annex III and to prepare regular reports on the evaluation and

review of the regulations. Final provisions also provide a timeline on differentiated transitional period for the initial date of the applicability of the regulation.

Next steps

Following the EC's proposal in April 2021, the regulation could enter into force in the second half of 2022 with a transitional period. In this period, standards would be mandated and developed, and the governance structures set up would be operational. The second half of 2024 is the earliest time the regulation could become applicable to operators with the standards ready and the first conformity assessments carried out.

Other legislative proposals

The proposed AI Regulation is not a stand-alone piece of legislation. It must be read in tandem with other EC legislative proposals, such as:

(i) the draft Digital Services Act (with provisions on recommenders and research data access);
(ii) the draft Digital Markets Act (with provisions on AI-relevant hardware, operating systems and software distribution);
(iii) the draft Machinery Regulation (revising the Machinery Directive in relation to AI, health and safety, and machinery);
(iv) announced product liability revision relating to AI;
(v) the draft Data Governance Act (concerning data sharing frameworks).

In what follows, we will briefly touch upon selected provisions of some of these legislative initiatives.

4.2 Digital Services Act Package

4.2.1 Digital Services Act (DSA)

On 15 December 2020, the EC published a regulation proposal for the Digital Services Act (DSA Proposal).[296] The Proposal aims to harmonize rules on the provision of intermediary services in the internal market. The goal is to create a safer digital space where fundamental rights and European values are respected and protected while ensuring a level playing field to foster innovation, growth and competitiveness. The text revises the 20 years old E-commerce Directive and creates several layers of due diligence obligations for intermediary services. The DSA

[296] European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, OJ COM/2020/825 final, 15.12.2020. The text is only a proposal at this stage and pursue its democratic legislative path in the other EU co-legislators which are the European Parliament and the Council of the European Union. Therefore, the provisions may be subject to amendments or deletion.

distinguishes between four levels of actors, all based on the definition of 'information society service':

(i) providers of intermediary services (defined in Art. 2(f));
(ii) hosts and online platforms providers (defined in Art. 2(h));
(iii) online platforms; and
(iv) very large online platforms (VLOPS), (defined in Art. 25).

The proposal is asymmetric, which means that different obligations are created for the various providers targeted by the proposal (i.e. intermediary services, hosting services, online platforms and very large platforms). The goal is to match their role, size and impact in the online ecosystem with their obligations. In practice, obligations for the intermediary services providers apply to all the sub-categories. However, if the provider falls in one or several subcategories (hosting services, online platforms and very large platforms), it will have one, two or three extra layers of obligations applicable for his intermediary services.

Concretely, the DSA proposes a series of EU-wide obligations for digital services such as:

(i) Rules for the removal of illegal goods, services or content online (notice and action and obligations to provide information to users);
(ii) Requirements on terms of service;
(iii) New obligations for very large platforms to take risk-based action to prevent abuse of their systems;
(iv) Transparency measures, including on online advertising and recommender systems;
(v) Provisions on the access by researchers to key platform data;
(vi) Rules on traceability of business users in online market places;
(vii) Oversight structure, a new European Board for Digital Services, enhanced supervision and enforcement by the Commission of the very large platforms.

The DSA proposal considers the impact of the use of AI based tools used in online media. The preamble of the proposal underlines how algorithmic systems shape information flows online (e.g. via content prioritization, advertisement display and targeting or content moderation). They also create or may reinforce existing discrimination in content moderation. It further points to the need expressed by civil society and academics for algorithmic accountability and transparency audits, especially about how information is prioritized and targeted to users.

The detailed assessment of the abovementioned obligations falls outside the scope of this Deliverable and will be tackled in Deliverables D6.2 "Report on Policy for Content Moderation" and D7.3 "From platform liability to platform responsibility - analysis of the shifting policy approach, guidelines for the AI-On-Demand-Platform and policy recommendations". However, below we provide a brief assessment of those parts of the proposal which interplay with the development of the AI systems.

Transparency obligations for the use of AI in content moderation

According to the DSA proposal, providers of intermediary services must include in their terms and conditions, in a clear and ambiguous language, information on any policies, procedures, measures and tools used for the purpose of content moderation, including "algorithmic decision-making" and human review.[297] It is rather unclear at this point, what providing information on "any policies, procedures, measures and tools" might look like in practice.

Providers of hosting services shall put mechanisms in place to allow any individual or entity to notify them of the presence on their service of specific items of information that the individual or entity considers to be illegal content. Interestingly, Article 14(6) indicates that providers of hosting services might make use of automated means to make decisions about the notices. When confirming receipt of the notification of a notice they must provide information on such use.[298] When communicating their decision to remove or disable access to specific items of information provided by the uploader, irrespective of whether the means used for detecting, identifying or removing or disabling access to that information were automated or not, hosting providers must inform the recipient of the decision and provide a clear and specific statement of reasons for that decision. Article 15 lists the information which must be included in such a statement. According to the DSA Proposal, online platforms must also put in place an internal complaint-handling system for managing the complaints against a decision taken against information provided/uploaded by a recipient of their services. The decision on the complaint must not be solely taken based on automated means. The online platforms also have additional obligation when it comes to transparency reporting as they must include in their yearly report on content moderation, any use made of automatic means for the purpose of content moderation, including a specification of the precise purposes, indicators of the accuracy of the automated means in fulfilling those purposes and any safeguards applied.

Transparency obligations for online platforms displaying advertising

The online platforms displaying advertising on their online interfaces must also ensure that information on the ad is provided in a clear and unambiguous manner and in real time: such as on whose behalf the ad is displayed and meaningful information on the main parameters used to determine the recipient to whom the advertisement is displayed, in other words what are the criteria used for the targeting.[299]

VLOPS, for their parts, must conduct at least once a year a risks assessment of the functioning and use of their service, which must consider how their content moderation systems, recommender systems and systems for selecting and displaying advertisement influence any of the systemic risks.[300] Once risks are identified, mitigation measures must be put in place and can

[297] Art. 12 of the DSA Proposal.
[298] Art. 14 of the DSA Proposal.
[299] Art. 23 of the DSA Proposal.
[300] Art. 26 of the DSA Proposal.

lead to adapting content moderation or recommender systems, their decision-making processes, etc.

Transparency obligations for recommender systems

The DSA Proposal defines a "recommender system" as "a fully or partially automated system used by an online platform to suggest in its online interface specific information to recipients of the service, including as a result of a search initiated by the recipient or otherwise determining the relative order or prominence of information displayed".[301] The Proposal notes that such recommender systems can have a significant impact on the ability of recipients to retrieve and interact with information online. They also play an important role in the amplification of certain messages, the viral dissemination of (mis)information and the stimulation of online behavior.[302] Consequently, the VLOPS shall set out in their terms and conditions, in a clear, accessible, and easily comprehensible manner, the main parameters used in their recommender systems, as well as any options for the recipients of the service to modify or influence those main parameters that they may have made available, including at least one option which is not based on profiling, within the meaning of the GDPR.[303] This provision has already sparked some debate. Most notably, the EDPS notes that recommender systems should by default not be based on "profiling" and that in line with the GDPR requirements of the data protection by design and by default and the data minimization principle, recommender systems should be based on opt-in rather than opt-out.[304]

Moreover, VLOPS must also empower recipients to modify their recommender systems by providing an easily accessible functionality on their interface allowing the recipient of the service to select and modify the parameter on which the recommender system determines the relative order of information presented.[305]

Access to data

The DSA Proposal also provides a specific provision on data access and scrutiny. Article 31 requires VLOPS to provide the Digital Services Coordinator of the Member State where the provider of an intermediary service is established or its legal representative resides or is established, or the EC, upon their reasoned request and within a reasonable period, access to data that are necessary to monitor and assess compliance with the DSA. That Digital Services Coordinator and the EC shall only use that data for those purposes. Under certain conditions,

[301] Art. 2(o) of the DSA Proposal.
[302] Recital 62 of the DSA Proposal.
[303] Art. 29 of the DSA Proposal.
[304] 'EDPS, Opinion 1/2021 on the Proposal for a Digital Services Act', 10 February 2021, < https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf> .
[305] Art. 29(2) of the DSA Proposal.

VLOPS will be obliged to provide within a reasonable period, access to data to "vetted researchers".[306] For detailed assessment of this provision see Section 5 below.

Access and explanations on the algorithms used

During on-site inspections, the EC and auditors or experts appointed by it may require VLOPS to provide explanations on its organisation, functioning, IT system, algorithms, data-handling and business conducts.[307] Moreover, in order to enhance Supervision, investigation, enforcement and monitoring of VLOPS' activities, the DSA provides that the EC should be empowered to require access to and explanations relating to databases and algorithms of relevant persons, and to interview, with their consent, any persons who may be in possession of useful information and to record the statements made.[308]

4.2.2 The Digital Markets Act (DMA)

The proposal for regulation on contestable and fair markets in the digital sector better known as the Digital Markets Act (DMA), was released on 15 of December 2020. The text introduces rules for platforms acting as gatekeepers to prevent them from unfair practices such as imposing unfair conditions on businesses and consumers and ensuring the openness of important digital services. The text aims to complement the competition framework by adding specific rules for the digital actors. The DMA provisions are without prejudice of the existing competition rules both European (such as article 101, 102 TFEU) and the national competition rules applicable regarding unilateral behavior.

Gatekeepers are defined by Article 3 of the proposal and need to meet three cumulative conditions to fall into the scope of the obligations of this proposal. A provider of core platform services shall be designated as gatekeeper if it has a significant impact on the internal market; it operates a core platform service which serves as an important gateway for business users to reach end users; and it enjoys an entrenched and durable position in its operations or it is foreseeable that it will enjoy such a position in the near future.

Part 2 of the impact assessment report contains the unfair practices which were listed by the respondents to the Open Public Consultation on the Digital Services Act package.[309] Two of the unfair practices listed are linked with the use of AI and algorithms. Firstly, the report explains how self-preferencing practices and the dual role of gatekeepers threatens fair competition and consumer welfare. Self-preferencing is considered to be a quite common practice deployed by large, vertically integrated platforms. Responses by business users suggest that search and

[306] Art. 31(2) of the DSA Proposal.
[307] Art. 54(3) of the DSA Proposal.
[308] Art. 57 of the DSA Proposal.
[309] European Commission, Commission Staff Working Document Impact Assessment Report accompanying the proposal for a Digital Markets Acts, SWD(2020) 363 final, Part 2/2, 15.12.2020.

ranking algorithms often give preference to the platform's own services.[310] Secondly, the general lack of transparency on business practices on platforms (e.g. use of algorithms, content prioritization) has been underlined as an unfair practice.[311] The information asymmetry between the gatekeepers and the rest of the actors was also pointed out as a prominent issue. These providers benefit from large amounts of data and analytics gathered by the use of their services by business. The use of advanced algorithms and machine learning techniques by gatekeepers facilitates targeting, discriminatory practices, and behavioral manipulation.[312]

The report highlighted how the issue of algorithmic transparency and accountability was crucial for solving the unfair practices issues of gatekeepers and the prohibition of discrimination through self-preferencing was also underlined.[313]

On media pluralism, the impact assessment report relays the concern regarding the fair remuneration of press publishers for their content used by or uploaded on large online platforms. The report also points that consumers should be informed about the rationale behind the choice of the content displayed.[314]

To mitigate the abovementioned concerns, once a company is considered a 'gatekeeper', the DMA sets out obligations and prohibitions for its core platform services. Art. 6 of the DMA Proposal prevents the gatekeeper from, *inter alia*: (i) combining personal data from their core platform services with data not publicly available from other sources (including other services offered by the gatekeeper); (ii) treating more favourable in ranking services and products offered by the gatekeeper itself; (iii) technically restricting the ability of end users to switch between different software applications.

Gatekeepers must also comply with the set of obligations, in particular: (i) they must allow the installation and use of third-party software or applications which are interoperable with the gatekeeper's core platform services; (ii) must provide business users and third parties with access to data, including consumer data, '*provided for or generated in the context of*' their use of core platform services. Similarly to the AI Act proposal, transparency regarding the use of AI systems and algorithms is an approach which has also been followed by the DMA. When it comes to investigation, enforcement and implementation of the DMA, Art. 19 of the proposal provides that the EC may also request access to databases and algorithms of undertakings and request explanations on those by a simple request or by a decision. On-site inspections are also foreseen and the EC and auditors or experts appointed by it may require gatekeepers to provide access to algorithms. Fines are expected in case of non-compliance with these obligations. The Proposal

[310] European Commission, Commission Staff Working Document Impact Assessment Report accompanying the proposal for a Digital Markets Acts, SWD(2020) 363 final, Part 1/2, 15.12.2020. p. 11-12.
[311] ibid., p. 31.
[312] Ibid., p. 47.
[313] ibid, p. 39.
[314] ibid, p. 40.

also provides periodic penalty payments which can go up to 5% of the average daily turnover to compel the undertakings including the gatekeepers to ensure access to databases and algorithms and submit to an on-site inspection.

4.3 Data Governance Act (DGA)

In November 2020, the EC adopted the Proposal for a Data Governance Act (DGA proposal).[315] It is its first legislative initiative under the 2020 European Data Strategy that aims to reinforce the single market for data. The objective of the DGA proposal is to set the conditions for enhancing the development of the common European data spaces, by bringing trust in data sharing and data intermediaries. As provided by the Impact Assessment, although an increasing amount of data is being generated through the use of digital devices and services, the cross-border availability of such data for EU companies and researchers remains too limited. As a result, the EC fears that the European economy will increasingly depend on third countries, in particular for the development of the Internet of Things and AI systems.[316]

The EC identifies that the lack of a proper legal environment, as well as technical problems that hinder data sharing, such as the lack of cross-sectoral interoperability, the limited ability to obtain reusable data, and the uncertainty about data quality, impedes activities of the organisations conducting data-sensitive activities.[317] That's why, with the DGA Proposal, the EC wants to foster data sharing across the European digital single market. The DGA proposal consists of the three main pillars. First, it creates a compulsory notification regime for a range of data sharing services. Second, it introduces a voluntary registration regime applying to data altruism services. Third, the DGA proposal creates a legal regime for the re-use of public sector data which are subject to the rights of third parties.[318]

4.4 Data Act

The DGA Proposal is also intertwined with the forthcoming Data Act. In May 2021, the Commission published its Inception Impact Assessments on the forthcoming Data Act.[319] It provides that the Data Act initiative will aim to increase access to and further use of data, so that more public and private actors can benefit from techniques such as Big Data and machine learning. Currently, private contracts and 'big tech' companies often regulate the conditions of access and further usage in B2B relationships. The initiative would look both at data usage rights in industrial value chains to allow all parties to benefit from data and data-driven innovation.

[315] European Commission, Proposal for a Regulation of the European Parliament and the Council on European Data Governance (Data Governance Act), COM(2020) 767 Final.
[316] European Commission, Staff Working Document, Impact Assessment Report Accompanying the DGA Proposal, SWD(2020) 295 Final.
[317] ibid.
[318] Baloup J. and others, 'White Paper on the Data Governance Act' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3872703> accessed 16 August 2021.
[319] European Commission, Data Act - Inception Impact Assessment.

More specifically, the initiative seeks to address the following issues: (i) use of privately-held data by the public sector; (ii) fairness in B2B data sharing contracts to further facilitate access to data and data sharing; (iii) legal certainty on access and use of co-generated non-personal data, including data generated from the Internet of Things (IoT); (iv) safeguards for non-personal data in international contexts; and (v) establishing more competitive markets for cloud computing services and improving data and application portability between cloud computing services in the whole data economy.

At the same time, the Inception Impact Assessment foresees the review of the Database Directive with the aim to ensure that the application of the Database Directive, in particular the sui generis right, does not pose an obstacle to the access and use of machine-generated data and facilitate the sharing of such data.

At the time of writing, the EC is gathering the views on the Data Act until 3 September 2021. It is foreseen that the proposal for the Data Act will be tabled in Q3-Q4/2021.

5 The potential impact of the anticipated EU regulatory initiatives in the field of AI for the AI4Media project

There is no doubt that various anticipated and forthcoming EU policy and regulatory initiatives will have a profound impact both on research activities within the AI4Media project, as well as on the commercial and non-commercial activities undertaken by AI4Media partners. In the following paragraphs, we briefly discuss the potential impact of the EU regulatory initiatives in the field of AI for the AI4Media project.

Data and data access for researchers

The availability of social media data for academic research and journalism is a major challenge. The social media companies have no incentive and no interest in revealing what kind of data they have on users and how this data is being used. Some social media data is accessible through Application Programming Interfaces (APIs), but most of the major social media companies are making it difficult for academics and journalists to obtain comprehensive access to their data.[320] Access to social media platforms' data for researchers is currently mainly governed by contractual agreements and platforms' own terms of service. As provided in the Assessment of the Code of Practice on Disinformation, "it is a shared opinion amongst European researchers that the provision of data and search tools required to detect and analyse disinformation cases is still episodic and arbitrary and does not respond to the full range of research needs."[321]

The most recent example is Facebook's shut down of the accounts of researchers using access to Facebook Ad Library to study political advertising and misinformation within the Ad Observatory at New York University (NYU).[322] In Facebook's view, the NYU's Ad Observatory project studied political ads using unauthorized scraping to access and collect data from Facebook, in violation of the platform's terms of service.[323]

It is clear that as provided by the "Artificial intelligence in the audio-visual sector" report by the European Audio-visual Observatory, "we need more data and independent research on the availability of different types of content, the consumption and engagement with that content, the participants involved in this process and the impact of these processes on individual and

[320] Batrinca B. and Treleaven P.C., 'Social Media Analytics: A Survey of Techniques, Tools and Platforms' (2015) 30 AI & SOCIETY 89.
[321] European Commission, The Staff Working Document (SWD (2020)180 Final - Assessment of the Code of Practice on Disinformation).
[322] Hatmaker T, 'Facebook cuts off NYU researcher access, prompting rebuke from lawmakers' <https://techcrunch.com/2021/08/04/facebook-ad-observatory-nyu-researchers/> accessed 20 August 2021.
[323] Clark M, 'Research Cannot Be the Justification for Compromising People's Privacy', <https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/> accessed 15 August 2021.

collective democratic and cultural performances".[324] Recent regulatory initiatives, such as the Digital Services Act (see Section 4.2. above), try to address this problem. Article 31 of the DSA proposal provides a specific provision on data access and scrutiny. It imposes an obligation on the very large online platforms to provide the Digital Services Coordinator of establishment or the Commission, upon their reasoned request and within a reasonable period, access to data that are necessary to monitor and assess compliance with the DSA. Upon a reasoned request from the Digital Services Coordinator of establishment or the Commission, VLOPS shall also provide access to data to "vetted researchers". According to Art. 31(4), "in order to be vetted, researchers shall be affiliated with academic institutions, be independent from commercial interests, have proven records of expertise in the fields related to the risks investigated or related research methodologies, and shall commit and be in a capacity to preserve the specific data security and confidentiality requirements corresponding to each request." The reasoned request to access data must, however, come from the Digital Services Coordinator of establishment or the Commission. Moreover, according to EU DisinfoLab, the new rules pose overly restrictive criteria needed for "vetted researchers", narrowing the scope to university academics. This is not likely to facilitate access to data to a variety of different actors: journalists, educators, web developers, fact-checkers, digital forensics experts, and open-source investigators.[325] The final scope of this provision will, undoubtedly, shape the way in which (vetted) researchers, journalists, and social activist will be able to access platforms' data. This is particularly relevant for the AI4Media WP6 activities such as opinion mining and automated extraction of public opinion from social media platforms such as Twitter (Task T6.4 – AI for Healthier Political Debate) that currently rely on the APIs.

While discussing access to (media) data, it is worth mentioning that in December 2020, the Commission adopted an Action Plan to support the recovery and transformation of the media and audio-visual sector (Media and Audio-visual Action Plan).[326] The Media and Audio-visual Action Plan aims to support the recovery and transformation of the media and audiovisual sector. It addresses the financial viability of the media sector to help the media industry recover and fully seize the opportunity of digital transformation, and further support media pluralism. Importantly, under Action 4 'Unleashing innovation through a European media data space and encouraging new business models', the EC proposes the concept of the "media data space" to support media companies in sharing data and developing innovative solutions.

---

[324] Cappello M (ed.), 'Artificial intelligence in the audiovisual sector', IRIS Special, European Audiovisual Observatory, Strasbourg, 2020 <https://rm.coe.int/iris-special-2-2020en-artificial-intelligence-in-the-audiovisual-secto/1680a11e0b>

[325] EU DisinfoLab's contribution to the Commission's second call for feedback on Digital Services Act, <https://www.disinfo.eu/advocacy/how-the-digital-services-act-%28dsa%29-can-tackle-disinformation/>, accessed 15 August 2021.

[326] European Commission, Communication from the Commission, Europe's Media in the Digital Decade: An Action Plan to Support Recovery and Transformation COM/2020/784 final.

The EC recognizes the importance of data for the media sector: "data spaces can change the way in which creators, producers, and distributors collaborate. They host relevant media data such as content, audience data and content meta-data as well as other types of data on users' behaviors that might be useful to create content better tailored to consumer needs and distribute it more efficiently."[327] The initiative of a European "media data space" builds on the European Data Strategy and the proposed Data Governance Act (see Section 4.3.). While it remains to be seen what shape the media data space will take, the creation of a shared data space will facilitate European fact-checking networks in news verification and help fact-checkers to have access to the relevant data to the spread of disinformation.

Finally, it is worth reminding that there are growing data and data governance requirements while training AI systems. As explained above (Section 4.1.3.), Art. 10 of the AI Act provides that high-risk AI systems which make use of techniques involving the training of models with data shall be developed on the basis of training, validation and testing data sets that meet the quality criteria such as appropriate data governance and management practices. Also, training, validation and testing data sets shall be relevant, representative, free of errors and complete. If adopted, these legally binding requirements will set a high standard on processing data.

Although, in the current draft AI Act, this provision applies only to 'high-risk' AI systems, the scope of this provision may be changed before the final text is adopted.

One the other hand, one must keep in mind that privacy and data governance requirement of the AI HLEG (see Section 3.2.2.1 above) applies to all AI systems, regardless of the context (high risk or low-risk, academic research or commercial application). Similarly, the GDPR already contains legally binding requirements on all data processing activities (including in the context of AI systems) which involve personal data, such as the obligation of a lawful ground for data processing activities and adherence to GDPR principles.

Academic research exception in the AI Act

The key issue which comes to the fore is the scope of exceptions for academic research. It is important to note that a research exception is only provided in a recital, it is not dealt with elsewhere in the text of the proposed Regulation. Recital 16 of the AI Act deals with the prohibition of placing on the market, putting into service or use of certain AI systems intended to distort human behaviour, whereby physical or psychological harms are likely to occur. Such AI systems deploy subliminal components individuals cannot perceive or exploit vulnerabilities of children and people due to their age, physical or mental incapacities. They do so with the intention of materially distorting the behaviour of a person and in a manner that causes or is likely to cause harm to that or another person. Research for legitimate purposes in relation to such AI systems, however, should not be stifled by the prohibition, if such research does not

[327] ibid.

amount to use of the AI system in human-machine relations that exposes natural persons to harm and such research is carried out in accordance with recognized ethical standards for scientific research. Should this provision be conceived as a general exception for research or a special exception only related to prohibited AI practices referred to in recital 16 (i.e., systems intended to distort human behaviour, whereby physical or psychological harms are likely to occur), but not to other categories (e.g., biometric identification systems, social scoring systems)? It seems that this recital only addresses Art. 5(1) (a) and (b).

As a consequence, researchers will have to comply with other AI regulation obligations i.e. related to "high-risk systems" or certification procedures. Moreover, the recital provides that "research (…) should not be stifled", however, only if research is done "for legitimate purposes", "if such research does not amount to the use of AI (…) that exposes natural persons to harm" and "is carried out in accordance with recognised ethical standards". The meaning of these notions, especially the notion of "research for legitimate purposes" is unclear, which may adversely affect legal certainty of researchers.

AI Act's applicability to media applications

In addition, the scope of the AI act is not clear for the AI systems applied in the media sector, which might impact the research activities of the AI4Media project. For instance, it is not clear whether the provision which prohibits the use of subliminal techniques could cover some AI systems used in practice such as the recommender systems or systems used for targeted advertising. The requirements imposed on manipulative AI, such as the use of *subliminal* techniques or the exploitation of a specific vulnerability of a specific group of persons, as well as the requirement of intent, can result in these provisions having a limited scope. More incidental manipulative systems (such as targeted advertising) are therefore not likely to be covered.

Though the explanatory memorandum suggests that other existing instruments still cover manipulative or exploitative practices, apart from practices prohibited under Art. 5, it fails to address that none of this legislation explicitly contains provisions on manipulation. As Bublitz and Douglas emphasize, AI systems can powerfully influence or weaken control over individuals' thoughts and behaviours, by bypassing or weakening rational control.[328] This includes microtargeted advertisement, as well as abuse of trust in recommender systems and their influence on decision-making. Thus, these practices should also be deemed to be manipulative, and they must be fairly addressed in the AI Act, instead of simply referring to other legislation. Perhaps, they could be classified as high-risk AI if they substantially influence thought or behaviour in ways that bypass or weaken rational control.

[328] Bublitz J C, Douglas T, Manipulative Influence via AI Systems and the EU Proposal for Regulation of Artificial Intelligence, 2021. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665640_en.

The AI high-risk systems listed in the annex to the AI Act do not contain media applications, but the media sector is directly concerned when it comes to transparency obligations both in the AI Act proposal and in the DSA proposal.

The scope of the AI Act is also unclear when it comes to transparency obligations applicable to bots, emotion recognition systems and deepfakes (Art. 52 of the AI Act). It is recommended to pay a particular attention to how the "emotion recognition system" definition and applicable transparency obligations for such systems change as the AI Act proposal goes down the legislative path. In particular, will 'sentiment analysis' (Task 6.4) fall under "emotion recognition system" definition? Or, can measuring and predicting the user's affective response to multimedia content distributed on social media with the use of physiological signals (Task T6.6) be considered as such?

As a side note, according to the impact assessment of the AI Act, transparency obligations already exist in other cases which may involve AI such as when a person is subject to solely automated decisions or micro-targeted. The impact assessment reminds of the following legislation which provides transparency obligations: data protection legislation (Art. 13 and 14 of the GDPR), consumer protection law, the proposals for the e-Privacy Regulation and in the proposal for the Digital Services Act. However, as an example and as explained in Section 4.2.1, transparency provisions on recommender systems in Art. 29 DSA only apply to very large online platforms. Nevertheless, should the scope of this provision change in the upcoming DSA drafts, automatically ranking user profiles and recommending content (Task 6.7) can be subject to the new obligations. Similarly, provisions on targeted advertising in Art. 24 DSA only apply to online platforms, which is currently limiting the reach of such provision.

This shows that the AI Act proposal does not exist in a legal vacuum and existing legislation is equally applicable to AI systems. The applicability of various legal frameworks to various AI systems and various types of the platforms, makes, however, the current legal picture puzzling.

Algorithmic copyright filtering

Concerning AI technologies to detect IP infringements, both the resolution and the action plan encourage utilizing filtering tools. Though such implementation could bring convenience in terms of detecting any infringement faster than a human review, it is important to note that AI technologies are not sophisticated enough at the moment to analyse the nuances of copyright protection. Such algorithmic filtering especially creates issues concerning the detection of copyright limitations and exceptions.[329]

[329] Samuelson P., Pushing Back on Stricter Copyright ISP Liability Rules. Michigan Technology Law Review, 2020. https://ssrn.com/abstract=3630700.

In the academic context, public domain work and non-exclusive/open licenses such as Creative Commons licenses are heavily used. However, such filtering technologies usually do not come equipped with an aggregated database of non-exclusive licensed and/or public domain works. Thus, without clarifying further the effect of implementation of such technologies on the aforementioned exceptions, academic and creative work could suffer a great deal.

Additionally, practices of such filtering tools have the potential to interfere with freedom of expression by removing legal content, violating the rights of access to knowledge and freedom to share. Unfortunately, the issue of over removal is especially gaining prominence in countries where notice and staydown regimes have also been trending, perhaps influenced by the EU's policy on incentivizing the implementation of such tools. Furthermore, when it comes to relying on automated decision-making (algorithmic filtering) concerning legality or illegality of such works, it is always important to have ex-ante human review mechanisms before removing content to avoid any violations of fundamental rights, as well as preventing bad faith takedown notices.

The liability regime for hosting content and platform responsibility for third party infringing and/or illegal content will be particularly relevant in WP7 "*Integration with AI-on-Demand platform*".

6 Conclusions

This deliverable provides an overview of the EU policy on AI and the forthcoming EC legislative proposal on AI regulation. By doing so, the aim is mainly to provide clear overview to the AI4Media consortium of existing and upcoming policy frameworks and an analysis of the ensuing principles and requirements. The deliverable demonstrated how numerous and various the EU policy initiatives are when it comes to AI systems. As observed, both sector transversal and sector specific initiatives were adopted in the EU within the last few years given the rapid development of AI technologies. All EU policy documents acknowledge the opportunities brought by AI as well as the risks to fundamental human rights and citizens' concerns.

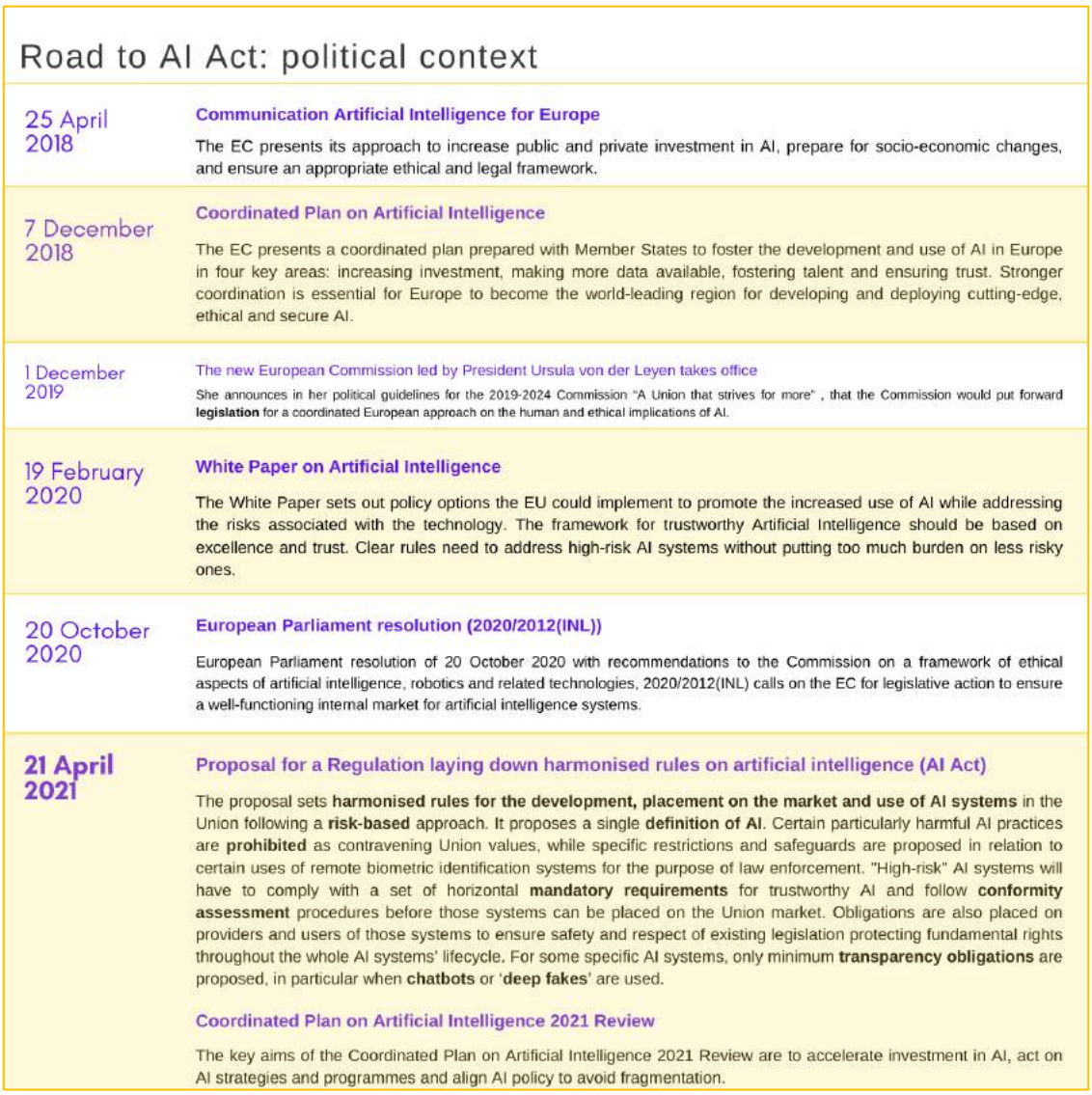

Figure 16: The Road to AI Act: political context

From the assessment of the overarching AI political initiatives (see Section 3.3.1), we could observe the trend calling for regulation and a switch from policy to legal initiatives. A chronological representation of this trend is shown in Figure 16 above.

This deliverable presented and analysed EU initiatives on AI ethics, intellectual property rights and safety and liability initiatives. The main AI ethics initiative at the EU level are the AI HLEG "Ethics Guidelines for Trustworthy Artificial Intelligence". Section 3.2.2 provided a detailed assessment of the guidelines' principles and requirements for trustworthy AI. The requirements are not to be seen as a mere theoretical concept but should be anchored in the AI system's design and architecture. Importantly, the proposed AI Act provides legally binding obligations for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security which derive from the Ethics Guidelines of the AI HLEG.

AI challenges the most traditional IP legal notions such as "copying", "originality", "creator", "author", or "inventiveness". How should the value of human creation be balanced against AI creation? Does the advent of AI require any changes to the existing IP frameworks? These are some of the questions which have been addressed by the EU initiatives on intellectual property rights and AI (see Section 3.2.3.) A summary of identified problems, proposed solutions and expected actions can be found in Figure 17.

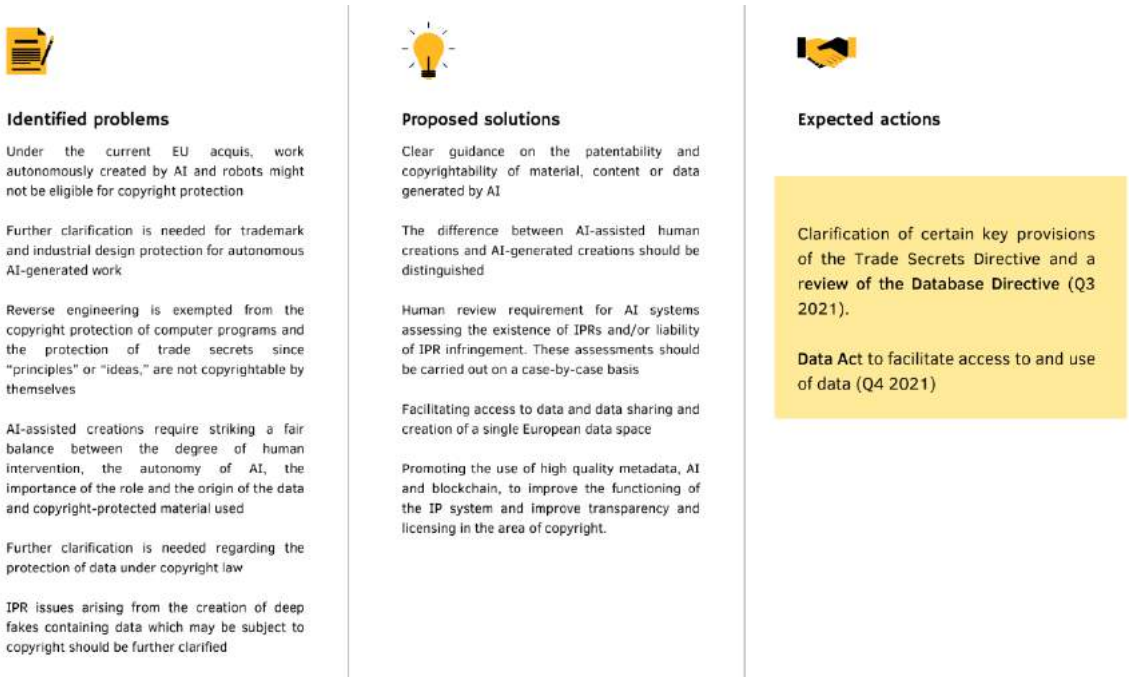

Figure 17: Intellectual property rights AI initiatives

Liability and safety are a recurring theme of the recent AI policy initiatives at the EU level. The essential characteristics of AI systems such as opacity, complex value chain and complex systems, autonomy, connectivity, and data dependency can make extremely hard for victims to obtain compensation with traditional rules not reflecting these specificities in legal modalities. Several options were put forward through reports and studies examined (see Section 3.2.4), but they all agreed that the current legal framework needs revisions. The selection of identified problems, proposed solutions and expected actions can be found in Figure 18.

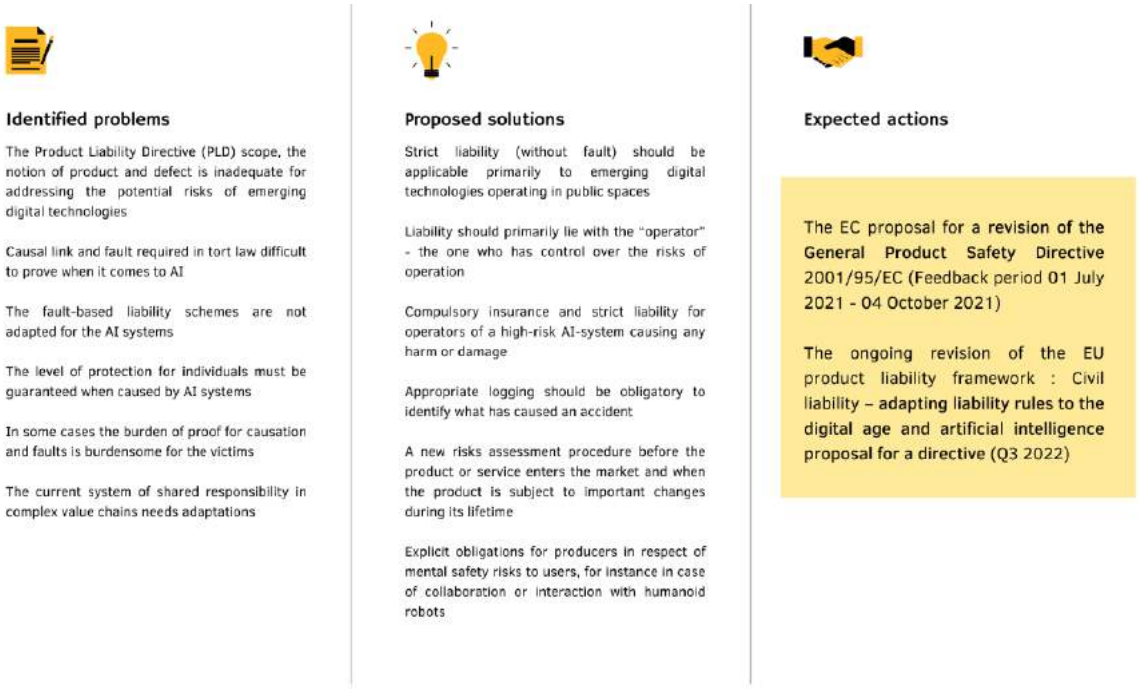

Figure 18: Safety and liability AI Initiatives

It is clear that there is a need for a coordinated strategy on developing a common European approach to trustworthy AI. All the analyzed documents point out the need to modernise and harmonise the current framework. To this end, we provided a comprehensive summary on new legislative proposals relevant for the AI4Media project containing provisions directly targeting AI systems (Section 4). Finally, we offered insights into how various anticipated and forthcoming EU policy and regulatory initiatives impact both research activities within the AI4Media project, as well as on the commercial and non-commercial activities undertaken by AI4Media partners. To that end, Section 5 aimed to anticipate this impact in four distinctive areas: (i) data and data access for researchers; (ii) academic research exception in the AI Act; (iii) AI Act's applicability to media applications; and (iv) algorithmic copyright filtering. The impact of these proposals on the AI4Media project is too early to determine with certainty. We will, however, closely follow the progress of the democratic legislative process. To this end, this deliverable will be followed by continued interactions with the consortium partners.

In order to provide both targeted guidance to partners and to be able to draw legal and regulatory conclusions from the interdisciplinary research in AI4Media, WP2 will extend this work in the context of Tasks 2.2, 2.3 and 2.4.

Considering the ever-changing legal landscape, this deliverable is not a one-off exercise. It provides a first step in our legal research and serves as a solid basis for the upcoming policy recommendations in the field of AI and Media (Task 2.2).

References

Legislation

Berne Convention for the Protection of Literary and Artistic Works (as amended on September 28, 1979)

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, *OJ L 130, 17.5.2019, p. 92–125*

Directive 2001/95/EC the European Parliament and of the Council of 3 December 2001 on general product safety, OJ L 11, 15.1.2002, p. 4–17.

Directive 2006/116/EC of the European Parliament and of the Council of 12 December 2006 on the term of protection of copyright and certain related rights, OJ L 372, 27.12.2006, p. 12–18

Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs, *OJ L 111, 5.5.2009, p. 16–22*

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases

European Commission, Proposal for a Regulation of the European Parliament and the Council on European Data Governance (Data Governance Act), COM(2020) 767 Final'.

European Patent Convention, 1973, available at: https://www.epo.org/law-practice/legal-texts/html/epc/1973/e/ma1.html

OJ L 77, 27.3.1996, p. 20–28 European Commission, COM(2021) 202 - Proposal for a Regulation of the European Parliament and of the Council on machinery products

Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93, OJ L 218, 13.8.2008, p. 30–47, ELI: http://data.europa.eu/eli/reg/2008/765/oj

Regulation (EC) No. 2008/765 setting out the requirements for accreditation and market surveillance relating to the marketing of products and Decision (EC) No. 2008/768 on a common framework for the marketing of products, OJ L218, 13.8.2008, p. 30–47

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), *OJ L 119, 4.5.2016, p. 1–88,* ELI: http://data.europa.eu/eli/reg/2016/679/oj

Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/201,1 OJ L 169, 25.6.2019, p. 1–44, ELI: http://data.europa.eu/eli/reg/2019/1020/oj

Treaty on European Union (TEU), *OJ C 326, 26.10.2012, p. 13–390,* ELI: http://data.europa.eu/eli/treaty/teu_2012/oj

CJEU case-law

CJEU C-05/08 Infopaq International v Danske Dagblades Forening (2009) ECLI:EU:C:2009:465 (Infopaq)

CJEU C-203/02 The British Horseracing Board Ltd and Others v William Hill Organisation Ltd. ECLI:EU:C:2004:695

CJEU C-310/17 Levola Hengelo BV v Smilde Foods BV (2018) ECLI:EU:C:2018:899

CJEU C-338/02 Fixtures Marketing Ltd v Svenska AB 2004 ECLI:EU:C:2004:696

CJEU C-444/02 Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE 2004 ECLI:EU:C:2004:697

CJEU C-46/02 Fixtures Marketing Ltd v Oy Veikkaus Ab 2004 ECLI:EU:C:2004:694

CJEU C-469/17 Funke Medien NRW GmbH v Bundesrepublik Deutschland (2019) ECLI:EU:C:2019:623 (Funke Medien)

CJEU C-683/17 Cofemel – Sociedade de Vestuário SA v G-Star Raw CV (2019) ECLI:EU:C:2019:721 (Cofemel)

CJEU Case C-145/10 Eva-Maria Painer v Standard VerlagsGmbH and Others 2013 ECLI:EU:C:2011:798

CJEU Case C-277/10 Martin Luksan v Petrus van der Let 2012 ECLI:EU:C:2012:65

CJEU Case C-572/13 Hewlett-Packard Belgium SPRL v Reprobel SCRL ECLI:EU:C:2015:750

Legal and policy documents

Bertolini A., Study on artificial intelligence and civil liability, (2020), https://www.europarl.europa.eu/RegData/etudes/STUD/2020/621926/IPOL_STU(2020)621926_EN.pdf

COMEST/UNESCO, “Report of COMEST on robotics ethics”, 2017. https://unescoblob.blob.core.windows.net/pdf/UploadCKEditor/REPORT%20OF%20COMEST%20ON%20ROBOTICS%20ETHICS%2014.09.17.pdf.

Council of Europe, ’AI initiatives‘, https://www.coe.int/en/web/artificial-intelligence/national-initiatives

Council of Europe, ’CAHAI Feasibility Study’ (17 December 2020), 1680a0c6da (coe.int)

Council of Europe, ’Compilation of contributions, Towards Regulation of AI systems – Global Perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe’s standards on human rights, democracy and the rule of law’ (December 2020), 1680a0c17a (coe.int)

Council of Europe, ’Conference of Ministers responsible for Media and Information Society Artificial intelligence – Intelligent politics Challenges and opportunities for media and democracy’ (2021), https://www.coe.int/en/web/freedom-expression/media2021nicosia

Council of Europe, ’Council of Europe's Work in progress‘, (29.06.2021 last update), https://www.coe.int/en/web/artificial-intelligence/work-in-progress

Council of Europe, 'Declaration by the Committee of Ministers on the manipulative capabilities of algorithmic processes' (2019), Result details (coe.int)

Council of Europe, 'Final Declaration of the Conference of Ministers responsible for Media and Information Society and resolutions on freedom of expression and digital technologies, on the safety of journalists, on the changing media and information environment, on the impacts of the COVID-19 pandemic on freedom of expression 10-11 June 2021', https://rm.coe.int/final-declaration-and-resolutions/1680a2c9ce.

Council of Europe, 'Guidance Note - Content Moderation, Best practices towards effective legal and procedural frameworks for self-regulatory and co-regulatory mechanisms of content moderation' (June 2021),1680a2cc18 (coe.int)

Council of Europe, 'Guidelines on Facial Recognition' (28 January 2021), 1680a134f3 (coe.int)

Council of Europe, 'Implications of AI-driven tools in the media for freedom of expression' (May 2020), 168097fa82 (coe.int)

Council of Europe, "European ethical charter on the use of artificial intelligence in judicial systems and their environment," 2019. https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c.

Council of the Europe, 'Recommendation CM/Rec(2020)1 of the Committee of Ministers to member States on the human rights impacts of algorithmic systems' (2020), https://search.coe.int/cm/pages/result_details.aspx?objectid=09000016809e1154

EDPB-EDPS Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act), 18 June 2021, < https://edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf>

EDPS, Opinion 1/2021 on the Proposal for a Digital Services Act

EDPS, Opinion 1/2021 on the Proposal for a Digital Services Act', 10 February 2021, < https://edps.europa.eu/system/files/2021-02/21-02-10-opinion_on_digital_services_act_en.pdf> .

EPO decision of 27 January 2020, EP 18 275 163. https://register.epo.org/application?number=EP18275163#_blank.

EPO, Guidelines for Examination, G-VI, 4. Enabling disclosure of a prior-art document.

EPO, Guidelines for Examination, Patentability Requirements.

Europa, Database Protection. https://europa.eu/youreurope/business/running-business/intellectual-property/database-protection/index_en.htm.

European Commission Expert Group on Liability and New Technologies, 'Report on liability for Artificial Intelligence and other emerging technologies' (2019), https://op.europa.eu/en/publication-detail/-/publication/1c5e30be-1197-11ea-8c1f-01aa75ed71a1/language-en.

European Commission, Communication of 25 November 2020, "Making the most of the EU's innovative potential, An intellectual property action plan to support the EU's recovery and resilience." https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52020DC0760.

European Commission, 'Expert Group on liability and new technologies', https://ec.europa.eu/transparency/expert-groups-register/screen/expert-groups/consult?do=groupDetail.groupDetail&groupID=3592&NewSearch=1&NewSearch=1

European Commission, “Gender Equality Strategy’’, 2020. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en.

European Commission, “Gender Equality Strategy’’, 2020. https://ec.europa.eu/info/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en.

European Commission, A Union that strives for more, My agenda for Europe : political guidelines for the next European Commission 2019-2024, < https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf>, accessed 15 July 2021.

European Commission, Commission Report of 19 February 2020, on safety and liability implications of AI, the Internet of Things and Robotics. https://ec.europa.eu/info/publications/commission-report-safety-and-liability-implications-ai-internet-things-and-robotics-0_en.

European Commission, Commission Staff Working Document - Impact Assessment accompanying the AI ACT proposal, SWD(2021) 84 final, 21.04.2021, PART 1/2.

European Commission, Commission Staff Working Document Impact Assessment Report accompanying the proposal for a Digital Markets Acts, SWD(2020) 363 final, Part 2/2, 15.12.2020.

European Commission, Commission Staff Working Document Impact Assessment Report accompanying the proposal for a Digital Markets Acts, SWD(2020) 363 final, Part 1/2, 15.12.2020.

European Commission, Communication from the Commission, Europe’s Media in the Digital Decade: An Action Plan to Support Recovery and Transformation COM/2020/784 final.

European Commission, Data Act - Inception Impact Assessment

European Commission, Directorate-General for Communications Networks, Content and Technology (European Commission), and others, “Trends and Developments in Artificial Intelligence: Challenges to the Intellectual Property Rights Framework: Final Report.” Publications Office of the European Union, 2020. https://data.europa.eu/doi/10.2759/458120.

European Commission, Directorate-General for Justice and Consumers (European Commission), and others Algorithmic Discrimination in Europe: Challenges and Opportunities for Gender Equality and Non Discrimination Law. Publications Office of the European Union, 2021.

European Commission, New rules for Artificial Intelligence – Facts page, <https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/excellence-trust-artificial-intelligence_en> accessed 10 June 2021.

European Commission, Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC, OJ COM/2020/825 final, 15.12.2020.

European Commission, Proposal for a Regulation of the European Parliament and of the Council on a Single Market For Digital Services (Digital Services Act) and amending Directive 2000/31/EC COM/2020/825 final

European Commission, Proposal for a Regulation of the European Parliament and of the Council on contestable and fair markets in the digital sector (Digital Markets Act) COM/2020/842 final

European Commission, Proposal for a regulation of the European Parliament and of the CouncIL on European data governance (Data Governance Act) COM/2020/767 final

European Commission, Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts COM/2021/206 final

European Commission, Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee on the Application of the Council Directive on the approximation of the laws, regulations, and administrative provisions of the Member States concerning liability for defective products (85/374/EEC), COM(2018) 246 final, https://eur-lex.europa.eu/legal-content/en/ALL/?uri=CELEX:52020DC0064 https://ec.europa.eu/transparency/documents-register/detail?ref=COM(2018)246&lang=en

European Commission, Report to the European Parliament, the Council and the European Economic and Social Committee, on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics, OJ, C(2020) 64 final, 19.02.2020,

European Commission, Staff Working Document, Impact Assessment Report Accompanying the DGA Proposal, SWD(2020) 295 Final

European Commission, Staff Working Document, Impact Assessment Report Accompanying the DGA Proposal, SWD(2020) 295 Final.

European Commission, Staff Working Document, Impact Assessment Report Accompanying the DGA Proposal, SWD(2020) 295 Final.

European Commission, The High Level Expert Group on AI, Ethics guidelines for trustworthy AI, 2019. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

European Commission, The Ministerial Declaration on eGovernment – the Tallinn Declaration, October 2017. https://ec.europa.eu/digital-single-market/en/news/ministerial-declaration-egovernment-tallinn-declaration).

European Commission, The Staff Working Document (SWD (2020)180 Final - Assessment of the Code of Practice on Disinformation).

European Commission, White Paper on Artificial Intelligence – a European approach to excellence and trust, OJ COM(2020) 65 final, 19.02.2020, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2020:65:FIN

European Commission. High Performance Computing Joint Undertaking; Research and Innovation. https://digital-strategy.ec.europa.eu/en/policies/high-performance-computing-joint-undertaking; https://ec.europa.eu/info/research-and-innovation_en.

European Council Meeting 14/17, October 2017. http://data.consilium.europa.eu/doc/document/ST-14-2017-INIT/en/pdf.

European Parliament, 'Draft Report with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL), https://www.europarl.europa.eu/doceo/document/JURI-PR-582443_EN.pdf?redirect

European Parliament, 'Resolution on automated decision-making processes: ensuring consumer protection and free movement of goods and services', 2019/2915, 12.02.2020 https://www.europarl.europa.eu/doceo/document/TA-9-2020-0032_EN.pdf

European Parliament, 'Study for the JURI Committee on European Civil Law Rules for Robotics', (2016), https://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf

European Parliament, Draft Report with recommendation to the Commissions on a Civil Liability regime for artificial intelligence, 2020/2014, 27.04.2020, https://www.europarl.europa.eu/doceo/document/JURI-PR-650556_EN.pdf

European Parliament, Resolution of 19 May 2021 on artificial intelligence in education, culture and the audiovisual sector (2020/2017(INI)), https://www.europarl.europa.eu/doceo/document/TA-9-2021-0238_EN.pdf

European Parliament, Resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)), TA (europa.eu)

European Parliament, the Resolution of 20 October 2020, on "Intellectual Property Rights for the Development of Artificial Intelligence Technologies (2020/2015(INI)." https://www.europarl.europa.eu/doceo/document/TA-9-2020-0277_EN.pdf.

European Parliament. Directorate General for Parliamentary Research Services., The Ethics of Artificial Intelligence: Issues and Initiatives. (Publications Office 2020) <https://data.europa.eu/doi/10.2861/6644> accessed 21 April 2021

European Parliamentary Research Service, Study on a common EU approach to liability rules and insurance for connected and autonomous vehicles (2018), https://www.europarl.europa.eu/RegData/etudes/STUD/2018/615635/EPRS_STU(2018)615635_EN.pdf

Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies, European Parliament, P9_TA(2020)0275'

Framework of Ethical Aspects of Artificial Intelligence, Robotics and Related Technologies, European Parliament, P9_TA(2020)0275'

High-Level Expert Group on AI, 'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment' (European Commission 2019)'

High-Level Expert Group on AI, 'Assessment List for Trustworthy Artificial Intelligence (ALTAI) for Self-Assessment' (European Commission 2019)'.

High-Level Expert Group on AI, "Ethics Guidelines for Trustworthy AI" (European Commission 2019)'

International Telecommunication Union (ITU), AI Repository, https://www.itu.int/en/ITU-T/AI/Pages/ai-repository.aspx

International Telecommunication Union (ITU), 'AI for Good', https://aiforgood.itu.int/

International Telecommunication Union (ITU), 'Journal on Future and Evolving Technologies', ITU Journal on Future and Evolving Technologies (ITU J-FET)

International Telecommunication Union (ITU), 'United Nations Activities on Artificial Intelligence 2020', https://www.itu.int/pub/S-GEN-UNACT-2020-1

Leaders of the G7, "Common vision for the future of artificial intelligence", 2018. https://www.mofa.go.jp/files/000373837.pdf.

Montreal AI Ethics Institute, 'State of AI Ethics Reports' (July 2021), https://montrealethics.ai/volume5/

OECD, 'Recommendation of the Council on Artificial Intelligence', (2019), OECD Legal Instruments

OECD, 'The OECD Artificial Intelligence Policy Observatory', https://www.oecd.ai/

OSCE Representative on Freedom of the Media, 'Policy paper on freedom of the media and artificial intelligence' (2020), 472488.pdf (osce.org)

OSCE, "#SAIFE: Presentation of spotlight initiatives", https://www.osce.org/fom/ai-free-speech/spotlight-initiatives

United Nations, 'Resource Guide on Artificial Intelligence Strategies' (2021), Resource Guide on AI Strategies April 2021_rev_0.pdf (un.org)

University of Montreal, "Montreal declaration for a responsible development of artificial Intelligence", 2017. https://www.montrealdeclaration-responsibleai.com/.

World Intellectual Property Organization, Understanding Copyright and Related Rights. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.

World Intellectual Property Organization, Understanding Copyright and Related Rights. https://www.wipo.int/edocs/pubdocs/en/wipo_pub_909_2016.pdf.

Academic resources

AI Now Institute, "The AI now report: the social and economic implications of artificial intelligence technologies in the near-Term", 2016. https://ainowinstitute.org/AI_Now_2016_Report.pdf.

Alan Turing Institute, 'A Primer Artificial Intelligence, Human Rights, Democracy and the Rule of law' (June 2021), 1680a2fd4a (coe.int)

Algo.Rules, "Rules for the design of algorithmic systems", 2019, https://algorules.org/en/home.

Article 19, "EU: Better Human Rights Protections Needed in HLEG Guidelines on AI.", 2019. https://www.article19.org/resources/eu-better-human-rights-protections-needed-in-hleg-guidelines-on-ai/.

Baloup J and others, 'White Paper on the Data Governance Act' [2021] SSRN Electronic Journal <https://www.ssrn.com/abstract=3872703> accessed 16 August 2021

Batrinca B and Treleaven PC, 'Social Media Analytics: A Survey of Techniques, Tools and Platforms' (2015) 30 AI & SOCIETY 89

Bibal A and Frénay B, 'Interpretability of Machine Learning Models and Representations: An Introduction' 7

Bibal A and others, 'Legal Requirements on Explainability in Machine Learning' (2021) 29 Artificial Intelligence and Law 149

Branch, 'AI and Climate Change: The Promise, the Perils and Pillars for Action', https://branch.climateaction.tech/issues/issue-1/ai-and-climate-change-the-promise-the-perils-and-pillars-for-action/

Brown S, Davidovic J and Hasan A, 'The Algorithm Audit: Scoring the Algorithms That Score Us' (2021) 8 Big Data & Society 205395172098386

Bryson JJ, 'The Artificial Intelligence of the Ethics of Artificial Intelligence: An Introductory Overview for Law and Regulation' 34

Bublitz J C, Douglas T, Manipulative Influence via AI Systems and the EU Proposal for Regulation of Artificial Intelligence, 2021. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/F2665640_en.

Buolamwini J and Gebru T, 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification' 15

Burk, Dan L., "Thirty-Six Views of Copyright Authorship", Houston Law Review, Vol. 58, 2020. https://ssrn.com/abstract=3570225.

Caplan R and others, 'Algorithmic Accountability: A Primer' (Data & Society 2018) <https://datasociety.net/output/algorithmic-accountability-a-primer/> accessed 12 February 2019

Cappello M (ed.), 'Artificial intelligence in the audiovisual sector', IRIS Special, European Audiovisual Observatory, Strasbourg, 2020 <https://rm.coe.int/iris-special-2-2020en-artificial-intelligence-in-the-audiovisual-secto/1680a11e0b>

Center for Democracy and Technology, "CDT's Comments to European Commission on Artificial Intelligence (AI HLEG)'s Draft Ethics Guidelines for Trustworthy AI.", 2020/ https://cdt.org/insights/cdts-comments-to-european-commission-on-artificial-intelligence-ai-hlegs-draft-ethics-guidelines-for-trustworthy-ai.

Cerna Collectif, "Research ethics in machine learning", 2018. https://hal.archives-ouvertes.fr/hal-01724307/document.

Chatila R and others, 'Trustworthy AI' in Bertrand Braunschweig and Malik Ghallab (eds), Reflections on Artificial Intelligence for Humanity, vol. 12600 (Springer International Publishing 2021) <http://link.springer.com/10.1007/978-3-030-69128-8_2> accessed 26 July 2021

Clark M, 'Research Cannot Be the Justification for Compromising People's Privacy', <https://about.fb.com/news/2021/08/research-cannot-be-the-justification-for-compromising-peoples-privacy/> accessed 15 August 2021

Edwards L and Veale M, 'Slave to the Algorithm? Why a "right to an Explanation" Is Probably Not the Remedy You Are Looking For' (LawArXiv 2017) preprint <https://osf.io/97upg> accessed 21 April 2021

Elkin-Koren, Niva and Perel (Filmar), Maayan, Algorithmic Governance by Online Intermediaries, Oxford Handbook of International Economic Governance and Market Regulation, 2018. https://ssrn.com/abstract=3213355

FATML, "Principles for accountable algorithms and a social impact statement for algorithms", 2016. https://www.fatml.org/resources/principles-for-accountable-algorithms.

Floridi L and others, 'AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations' (2018) 28 Minds and Machines 689

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... Schafer, B. (2018), "AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations", Minds and Machines, Vol. 28 No. 4, pp. 689-707.

Gilburt, B., "Women leading in AI: 10 principles of responsible AI", Towards Data Science, 2019. https://towardsdatascience.com/women-leading-in-ai-10-principles-for-responsibleai-8a167fc09b7d.

Ginsburg J C, "The Concept of Authorship in Comparative Copyright Law," 2014:

Ginsburg J C, The Concept of Authorship in Comparative Copyright Law, 52 DePaul L. Rev. 1063, 2003. https://via.library.depaul.edu/law-review/vol52/iss4/3

Goodman B and Flaxman S, 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"' [2016] arXiv:1606.08813 [cs, stat] <http://arxiv.org/abs/1606.08813>

Grgic-Hlaca, Nina, M. B. Zafar, K. Gummadi and Adrian Weller. "Beyond Distributive Fairness in Algorithmic Decision Making: Feature Selection for Procedurally Fair Learning", 2018.

Gupta A., Lanteigne C. and Kingsley S., (2020), 'SECure: A Social and Environmental Certificate for AI Systems', 'Computers and Society', https://arxiv.org/ftp/arxiv/papers/2006/2006.06217.pdf

Hamon R and others, 'Impossible Explanations?: Beyond Explainable AI in the GDPR from a COVID-19 Use Case Scenario', Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445917> accessed 24 May 2021

Hatmaker T, 'Facebook cuts off NYU researcher access, prompting rebuke from lawmakers', <https://techcrunch.com/2021/08/04/facebook-ad-observatory-nyu-researchers/> accessed 20 August 2021

Hildebrandt M, 'Privacy as Protection of the Incomputable Self: From Agnostic to Agonistic Machine Learning' (2019) 20 Theoretical Inquiries in Law 83

IEEE, "Ethically aligned design: a vision for prioritizing human well-being with autonomous and intelligent systems", Version 1, 2019. https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead1e.pdf.

International Conference of Data Protection and Privacy Commissioners (ICDPPC) (2018), "Declaration on ethics and data protection in artificial intelligence." http://globalprivacyassembly.org/wp-content/uploads/2018/10/20180922_ICDPPC-40th_AI-Declaration_ADOPTED.pdf.

Internet Society, "Artificial intelligence and machine learning: policy paper", 2017. https://www.internetsociety.org/resources/doc/2017/artificial-intelligence-and-machine-learning-policy-paper/.

ISP Liability Rules. Michigan Technology Law Review, 2020. https://ssrn.com/abstract=3630700

Kaminski M. E. and Malgieri G, 'Algorithmic Impact Assessments under the GDPR: Producing Multi-Layered Explanations' 29

Koshiama A. and Engin Z., 'Algorithmic Impact Assessment: Fairness, Robustness and Explainability in Automated Decision-Making', (2019), Data for Policy 2019: Digital Trust and Personal Data (Data for Policy 2019) (DFP), London. Zenodo, https://doi.org/10.5281/zenodo.3361708

Lacoste, A., Luccioni, A., Schmidt, V., & Dandres, T. (2019). Quantifying the Carbon Emissions of Machine Learning. ArXiv, abs/1910.09700.; Environmental Intelligence: Applications of AI to Climate Change, Sustainability, and Environmental Health (stanford.edu)

Latonero, M., "Governing artificial intelligence: upholding human rights and dignity", Data and Society, 2018.

M Ryan, BC Stahl, Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications, Journal of Information, Communication and Ethics in Society, 2020.

McKInsey Global Institute, 'Notes from the AI Frontier applying AI for social good, Discussion Paper' (2018) mgi-applying-ai-for-social-good-discussion-paper-dec-2018.pdf

Mireille van Eechoud, “Along the Road to Uniformity - Diverse Readings of the Court of Justice Judgments on Copyright Work,” JIPITEC 3, no. 1, 2012. http://www.jipitec.eu/issues/jipitec-3-1-2012/3322;

Mitrou L, ‘Data Protection, Artificial Intelligence and Cognitive Services: Is the General Data Protection Regulation (GDPR) “Artificial Intelligence-Proof”?’ [2018] SSRN Electronic Journal <https://www.ssrn.com/abstract=3386914> accessed 21 April 2021

Musikanski, L., Rakova, B., Bradbury, J. *and others* Artificial Intelligence and Community Well-being: A Proposal for an Emerging Area of Research. *Int. Journal of Com. WB* 3, 39–55 (2020). https://doi.org/10.1007/s42413-019-00054-6

NSTC, “The national artificial intelligence research and development strategic plan”, 2016. https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf.

Orian D., ’EU report on AI, new technologies and liability: key take-aways and limitations‘ (2020), https://www.law.kuleuven.be/citip/blog/eu-report-on-ai-new-technologies-and-liability-key-take-aways-and-limitations/

Pasquale F, The Black Box Society: The Secret Algorithms That Control Money and Information (Harvard University Press 2015)

Raji I.D., and others, ‘Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing’ [2020] arXiv:2001.00964 [cs] <http://arxiv.org/abs/2001.00964> accessed 27 July 2021

Raji ID and others, ‘Closing the AI Accountability Gap: Defining an End-to-End Framework for Internal Algorithmic Auditing’ [2020] arXiv:2001.00973 [cs] <http://arxiv.org/abs/2001.00973> accessed 11 August 2021

Renda A, ‘Europe: Toward a Policy Framework for Trustworthy AI’ in Markus D Dubber, Frank Pasquale and Sunit Das (eds), Andrea Renda, The Oxford Handbook of Ethics of AI (Oxford University Press 2020) <https://oxfordhandbooks.com/view/10.1093/oxfordhb/9780190067397.001.0001/oxfordhb-9780190067397-e-41> accessed 12 October 2020

Renda, Andrea. “Europe.” The Oxford Handbook of Ethics of AI, 1st ed., Oxford University Press, 2020.

Robbins S, ‘A Misdirected Principle with a Catch: Explicability for AI’ (2019) 29 Minds and Machines 495

Sage, “The ethics of code: Developing AI for business with five core principles”, 2017. https://www.sage.com/~/media/group/files/business-builders/business-builders-ethics-of-code.pdf.

Selbst A, ‘An Institutional View of Algorithmic Impact Assessments’ 35 78

Smart Dubai, “Artificial intelligence principles and ethics”, 2019. https://www.smartdubai.ae/docs/default-source/ai-principles-resources/ai-ethics.pdf.

Tatiana-Eleni Synodinou, The Foundations of the Concept of Work in European Copyright Law, in: Synodinou (ed.), Codification of European Copyright Law.

Tomašev, N., Cornebise, J., Hutter, F. *and others* AI for social good: unlocking the opportunity for positive impact. *Nat Commun* 11, 2468 (2020). https://doi.org/10.1038/s41467-020-15871-z

Veale M and Zuiderveen Borgesius F, ‘Demystifying the Draft EU Artificial Intelligence Act’ (SocArXiv 2021) preprint <https://osf.io/38p5f> accessed 20 July 2021

Vinuesa, R., Azizpour, H., Leite, I. and others, The role of artificial intelligence in achieving the Sustainable Development Goals. Nat Commun 11, 233 (2020). https://doi.org/10.1038/s41467-019-14108-y

Visser and others, Visser's Annotated European Patent Convention (EPC), 138.

Wachter S, Mittelstadt B and Floridi L, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76

Whittlestone J and others, 'The Role and Limits of Principles in AI Ethics: Towards a Focus on Tensions', Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society (ACM 2019) <https://dl.acm.org/doi/10.1145/3306618.3314289> accessed 12 October 2020

Other

Bedingfield W, The Wired, We finally know how bad for the environment your Netflix habit is, (2021), https://www.wired.co.uk/article/netflix-carbon-footprint

EU DisinfoLab's contribution to the Commission's second call for feedback on Digital Services Act, <https://www.disinfo.eu/advocacy/how-the-digital-services-act-%28dsa%29-can-tackle-disinformation/>, accessed 15 August 2021.

Hern A, The Guardian, 'Facebook and Google announce plans to become carbon neutral' (2020), https://www.theguardian.com/environment/2020/sep/15/facebook-and-google-announce-plans-become-carbon-neutral
Icelandic Institute for Intelligent Machines, "Ethics policy", 2015. https://www.iiim.is/ethics-policy/.

ICO (2017), "Big data, artificial intelligence, machine learning and data protection", 2017, https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf.

MIT Technology Review, 'Training a single AI model can emit as much carbon as five cars in their lifetimes' (2019),https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/

Rathenau Institute "Human rights in the robot age: Challenges arising from the use of robotics, artificial intelligence, and virtual and augmented reality", 2017. https://www.rathenau.nl/en/digitale-samenleving/human-rights-robot-age.

Responsible AI Licenses (RAIL), https://www.licenses.ai/

The Guardian, 'Facebook and Google announce plans to become carbon neutral' (2020), https://www.theguardian.com/environment/2020/sep/15/facebook-and-google-announce-plans-become-carbon-neutral

The Public Voice, AI Universal Guidelines, 2018. https://thepublicvoice.org/ai-universal-guidelines/.

Vincent J, The Verge, 'Twitter taught Microsoft's AI chatbot to be a racist asshole in less than a day' (2016), https://www.theverge.com/2016/3/24/11297050/tay-microsoft-chatbot-racist

Wiggers K, Venture Beat, Researchers propose framework to measure AI's social and environmental impact (2020), https://venturebeat.com/2020/06/12/researchers-propose-framework-to-measure-ais-social-and-environmental-impact.

Wild J, “Artificial Intelligence and the Future of the Patent System", I AM (Blog), 2018. https://www.iam-media.com/law-policy/artificial-intelligence-and-future-patent-system.

World Wide Web Foundation, “Artificial intelligence: Open questions about gender inclusion”, (2018). http://webfoundation.org/docs/2018/06/AI-Gender.pdf. European Commission, Communication Artificial Intelligence for Europe, April 2018. https://digital-strategy.ec.europa.eu/en/library/communication-artificial-intelligence-europe.

World Wide Web Foundation, “Artificial intelligence: Open questions about gender inclusion”, (2018). http://webfoundation.org/docs/2018/06/AI-Gender.pdf.

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