AI4media
ARTIFICIAL INTELLIGENCE FOR THE MEDIA AND SOCIETY

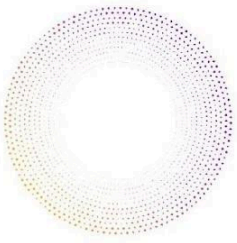
D4.4

Final analysis of the legal and ethical framework of trusted AI

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

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

info@ai4media.eu www.ai4media.eu



Deliverable title	Final analysis of the legal and ethical framework of trusted AI
Deliverable number	D4.4
Deliverable version	1.0
Previous version(s)	-
Contractual date of delivery	31 August 2023
Actual date of delivery	9 October 2023
Deliverable filename	D4.4 Final analysis of the legal and ethical framework of trusted AI_FINAL
Nature of deliverable	Report
Dissemination level	Public
Number of pages	83
Work Package	WP4
Task(s)	T4.1
Partner responsible	KUL
Author(s)	Lidia Dutkiewicz (KUL), Noémie Krack (KUL)* *This research is conducted under the supervision of Prof. Peggy Valcke and Dr Aleksandra Kuczerawy.
Editor	Lidia Dutkiewicz (KUL)
EC Project Officer	Evangelia Markidou

Abstract	Deliverable D4.4 provides the final analysis of the legal and ethical framework for trusted AI. After a reminder of data protection analysis from D4.3, the deliverable presents the various EU legislation updates impacting the data protection aspects of artificial intelligence. It then dives into specific themes relevant to the media sector through the lenses of their privacy and ethical considerations. These themes include fact- checking, recommender systems, and facial recognition. Finally, it focuses on the legal and ethical considerations of generative AI systems.

Keywords	Artificial Intelligence, Media, Ethical AI, ChatGPT, AI Act, GDPR, data protection, generative AI, fact-checking, recommender systems, facial recognition

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

Copyright

Contributors

NAME	ORGANISATION
Lidia Dutkiewicz	KUL
Noémie Krack	KUL

Peer Reviews

NAME	ORGANISATION
Anisa Halimi	IBM
Dan-Cristian Stanciu	UPB

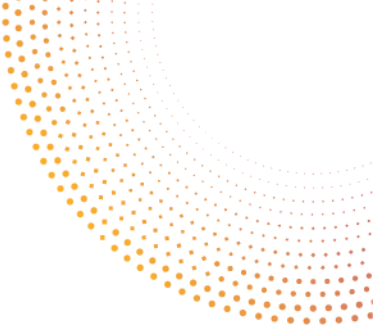
Revision History

VERSION	DATE	REVIEWER	MODIFICATIONS
0.1	24/02/2023	Lidia Dutkiewicz Noémie Krack	Initial table of contents
0.2	12/05/2023	Lidia Dutkiewicz Noémie Krack	Updated table of contents
0.3	28/07/2023	Lidia Dutkiewicz Noémie Krack	First draft
0.4	01/09/2023	Lidia Dutkiewicz Noémie Krack	Second draft
0.5	22/09/2023	Lidia Dutkiewicz Noémie Krack	Version ready for review
0.6	26/09/2023	Anisa Halimi	Internal review
0.7	5/10/2023	Dan-Cristian Stanciu	Internal review
0.8	06/10/2023	Lidia Dutkiewicz Noémie Krack	Reviews implementation
0.9	09/10/2023	Lidia Dutkiewicz	Final draft sent to CERTH
1.0	09/10/2023	Filareti Tsalakanidou	Final version for submission

Table of Abbreviations and Acronyms

Abbreviation	Meaning
AI	Artificial Intelligence
AI HLEG	High-level Expert on Artificial Intelligence
API	Application Programming Interface
Art.	Article
CCTV	Closed-Circuit Television
ChatGPT	Chat Generative Pre-trained Transformer
CJEU	Court of Justice of the European Union
Dir.	Directive
DGA	Data Governance Act
DMA	Digital Markets Act
DPA	Data protection authority
DPIA	Data Protection Impact Assessment
DSA	Digital Services Act
EC	European Commission
EDPB	European Data Protection Board
EDPS	European Data Protection Supervisor
EP	European Parliament
EPRS	European Parliamentary Research Service
EU	European Union
FAQ	Frequently Asked Questions
GDPR	General Data Protection Regulation
GPT	Generative Pre-trained Transformer
ICO	Information Commissioner's Office
ID	Identity
IP	Intellectual Property
LLM	Large Language Model
MEP	Member of the European Parliament


AI4media
ARTIFICIAL INTELLIGENCE FOR
THE MEDIA AND SOCIETY



Abbreviation	Meaning
ML	Machine Learning
NGO	Non-Governmental Organization
Rec.	Recital
TDM	Text and data mining
TFEU	Treaty on the functioning of the European Union
VLOP	Very Large Online Platform
VLOSE	Very Large Online Search Engine
WP29	Article 29 Data Protection Working Party
XAI	Explainable AI

Index of Contents

Index of Tables

Index of Figures

1 Executive Summary

Deliverable 4.4 provides the final analysis of the legal and ethical framework for trusted AI. This deliverable is a follow-up to D4.3 "*Initial analysis of the legal and ethical framework of trusted AI*". Having explained the basics of the General Data Protection Regulation (GDPR) applicability in the Artificial Intelligence (AI) context in D4.3, this deliverable goes more in-depth into the legal and ethical aspects of using (big) data in the AI media context.

Section 3 offers an update on EU legislation which has or may have an impact on existing data protection rules. It starts by recalling the key concepts and obligations of the GDPR, previously explained in the D4.3. Next, it provides an update on the AI Act proposal. Without going much in-depth into the AI Act, the deliverable focuses on the interplay between the GDPR and the AI Act by presenting the key amendments proposed by the European Parliament (EP). Then, the deliverable explains the links between the data protection legislation and the newly adopted legal acts, such as the Data Governance Act (DGA), the Digital Services Act (DSA), and the Data Act. It also offers an update on the guidelines coming from the European Data Protection Board (EDPB) and the European Data Protection Supervisor (EDPS), as regards the interpretation of the GDPR in the context of AI.

Section 4 contextualises the abovementioned findings by applying the EU data protection and ethical frameworks to a real-life scenario in the media sector. It does that through the lenses of three use cases to discuss the role that data protection and ethical consideration play in media AI applications: i) fact-checking on social media platforms, ii) recommender systems in public service media, and iii) use of facial recognition in broadcasting.

Section 5 focuses on the ethics of using big data, especially in the context of large generative AI models. It first dives into data protection considerations, by highlighting the most problematic compliance issues. To better illustrate the problem of compliance with the GDPR principles, the deliverable refers to a widely debated decision by the Italian Data Protection Authority (DPA) which ordered OpenAI to temporarily stop processing personal data. Next, the deliverable analyses the EP amendments to the AI Act proposal as regards the regulation of foundation models. The section explains the definition of foundation models, and related obligations for the providers of such models, including, but not limited to identifying and mitigating reasonably foreseeable risks, testing and evaluation, and drawing up documentation such as data sheets, model cards, and intelligible instructions for use. We also discuss the feasibility of compliance of the major foundation model providers with the AI Act requirements based on the research conducted by Stanford University. Next, Section 5 offers an overview of other legal considerations in developing and deploying large language models, such as in the field of Intellectual Property (IP) law, competition, and oversight. A set of ethical considerations is also considered. Section 6 offers conclusions.

2 Introduction

As the awareness of AI ethics has risen, both the public and private sectors have released numerous guidelines, frameworks, and ethical principles. Munn lists that there are over 50 relevant documents issued by government agencies, including national frameworks produced by the UK, the USA, Japan, China, India, Mexico, Australia, and New Zealand, amongst others. There are the Beijing AI Principles, DeepMind's Ethics, IEEE's Ethically Aligned Design, and many more.[1] The list of AI Principles at AI Ethicist now includes over 80 entries.[2] In the EU, the EC High- Level Expert Group released the Ethics Guidelines for Trustworthy Artificial Intelligence in 2019.

Since then, many have voiced the need for legally binding – instead of voluntary – principles that would guide the development of ethical AI. With the AI Act proposal released by the European Commission, we can observe a shift from a principle-based approach to AI Ethics towards legal requirements and obligations for the providers of AI applications. Together with enforcement, this can constitute a breakthrough in how AI technology is developed in the EU. However, since the AI Act is still being negotiated, many questions remain open. In this deliverable, we offer the latest updates on the AI Act, also regarding the development and use of foundation models and generative AI.

At the same time, the AI ethical discussions have since moved from high-level principles to more practical approaches and methodologies of embedded ethics. These have been discussed during two breaking sessions moderated by KUL researchers L. Dutkiewicz and N. Krack during the 2nd cross-cutting Theme Development Workshop on "Trusted AI - The future of creating ethical and responsible AI systems", co-organized by AI4Media on 13 of September 2023. [3] The main findings of these sessions also constitute part of this deliverable.

As regards the EU legal framework for trusted AI, the GDPR remains the core legislation that applies to AI data training, development, and deployment. Even though it does not explicitly mention artificial intelligence, it applies to any personal data processing, including those that take place in the data training phase of AI development. The AI Act, currently negotiated, should add more legal certainty for providers and users of AI systems, also in aspects that go beyond personal data processing.

Finally, recent developments in the field of large language models and generative AI bring new legal and ethical challenges for both developers and policy-makers. As we show in this deliverable some have called for a temporary pause on its developments, whereas others have suggested a need for robust EU regulation. The debate about which legal and ethical framework is needed is currently ongoing and its final shape will depend on the outcome of the AI Act

[1] Luke Munn, 'The Uselessness of AI Ethics' (2023) 3 AI and Ethics 869.

[2] https://www.aiethicist.org/ai-principles
[3] https://www.vision4ai.eu/tdw-trusted-ai/

negotiations. In this deliverable, we offer a view on the feasibility of compliance with the GDPR principles and with the likely AI Act obligations by generative AI.

3 Update on EU legislations

3.1 General Data Protection Regulation (GDPR)

For the sake of completeness, the following sections provide excerpts from Deliverable D4.3 to recall the key GDPR notions. For the full analysis of the applicability of the GDPR to AI, see Deliverable D4.3.

What is "personal data"?

Personal data is both data that make it possible to identify a natural person and data that relates to an identified or identifiable person. An individual is *identified* when, within a group of persons, they are distinguished from all other members of the group. Individuals can be identified by e.g., name or address ('direct identification'), but also by their IP address, cookie identifier, location data, or other factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that person ('indirect identification'). If a person cannot be immediately identified, it must be verified whether indirect identification is possible or not.

On the other hand, *identifiable* means that, although the person has not been identified yet, it is still possible to do so. To ascertain whether an individual is identifiable, Rec. 26 of the GDPR specifies that '*account should be taken of all the means reasonably likely to be used, such as singling out, either by the controller or by another person to identify the natural person directly or indirectly*'. Whether the means are 'reasonably likely' must be assessed in light of '*objective factors, such as the costs of and the amount of time required for identification, taking into consideration the available technology at the time of the processing and technological developments*'.

This means that establishing the identifiability of the person, and consequently the applicability of the GDPR, requires a dynamic, context-sensitive analysis of the factual situation. Thus, the same dataset might be considered as not containing personal data at the start of the processing and, later on, it might fall under the definition of 'personal data' given the tools and data available to the data controller. The same might happen depending on who is processing the datasets.

If the system processes personal data, then the GDPR must be complied with.

What are the special categories of personal data?

Special categories of personal data (also commonly called sensitive personal data) are:

- personal data revealing racial or ethnic origin;
- personal data revealing political opinions;
- personal data revealing religious or philosophical beliefs;
- personal data revealing trade union membership;
- genetic data;
- biometric data (where used for identification purposes);

- data concerning health;
- data concerning a person's sex life; and
- data concerning a person's sexual orientation.

You must always ensure that the data processing is generally lawful, fair, and transparent and complies with all the other principles and requirements of the GDPR. To process[4] personal data, you must always fulfil one of the lawful bases of Article 6 of the GDPR. In addition, you can only process special category data if you can meet one of the specific conditions in Article 9 of the GDPR.

What is the difference between anonymisation and pseudonymisation?

A major distinction has to be drawn between pseudonymized and anonymized data.

Pseudonymization is defined in Art. 4(5) of the GDPR as '*the processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person'.*

Anonymous information is defined in Rec. 26 of the GDPR as *"information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable".* In such case, the GDPR does not apply to the processing of such data.

The creation of a truly anonymized dataset from personal data whilst not depriving the information it carries from its added value is not a trivial task. Depending on the technical possibility and risks of re-identification, data may sometimes still be considered as personal, yet pseudonymized, from a legal perspective. The Article 29 Working Party (WP29, currently European Data Protection Board, EDPB) highlights that in determining whether or not the data are still identifiable, focus should be placed on the concrete means that would be necessary to reverse the anonymization technique, particularly the knowledge how to implement those means and the assessment of their likelihood and severity.[5] For example, encrypted personal data will be anonymous data, when it would require an excessively high effort or cost or it would cause serious disadvantages to reverse the process and re-identify the individual.

Additionally, one must bear in mind that the means to be assessed are not only those of the data controller, but also the ones that may be used by any other person (see Figure 1). **True**

[4] The term *'processing'* is very broad. It means any operation or set of operations performed on personal data or on sets of personal data, whether or not by automated means. It includes, but is not limited to: collection, recording, organisation, structuring, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, restriction, erasure or destruction.

[5] Article 29 Data Protection Working Party 2014, Opinion 05/2014 on Anonymisation Techniques.

anonymization is consequently a very onerous standard, and the notion calls for vigilance when used. In particular, having gone through anonymization process at a certain point in time should not be viewed as a silver bullet for circumventing the application of the GDPR, as identification of natural persons may happen in further processing activities (e.g., when aggregating such data with other data). In addition, critical views were expressed on the efficiency of anonymisation in a big data world, and researchers showed how even anonymised datasets can be traced back to individuals using machine learning.[6] They demonstrated that allowing data to be used to train AI algorithms would require much more work than simply adding noise, sampling datasets, and other de-identification techniques.

UPDATE

Importantly, on 26 April 2023, the General Court of the European Court of Justice (CJEU) issued a judgment in the case SRB v. EDPS (Case T-557/20) on the anonymisation of personal data. The General Court confirmed that personal views and opinions may indeed constitute personal data, but it must be investigated if a link to a particular person exists. The question at issue, was whether the data transmitted to the third party was related to an "identifiable" natural person. The court concluded that "it is not required that all the information enabling the identification of the data subject must be in the hands of one person". The General Court concluded that, to determine whether the information transmitted to the third party constitutes personal data, it is necessary to put oneself in the position of the third party and whether such a third party has the legal means to access the additional information to re- identify the data subject. The court focused on the recipient's view.

Figure 1 Update case SRB v. EDPS (Case T-557/20)

What are the different roles an organisation can play under the GDPR in an AI context?

One of the most important aspects under the GDPR is defining the different roles and responsibilities with regard to the processing of personal data. Throughout this deliverable we will be referring to **data controllers** and **data processors**. The distinction between (**data**) **controller** and (**data**) **processor** is important, because they each have different obligations under the GDPR.

At the different stages of the life cycle of an AI system, the **controller** is the natural or legal person, public authority or other organisation that **decides on the purposes and means** of processing personal data.

[6] Luc Rocher, Julien M Hendrickx and Yves-Alexandre De Montjoye, 'Estimating the Success of Re- Identifications in Incomplete Datasets Using Generative Models' (2019) 10 Nature Communications 3069.

Processor, on the other hand, means a natural or legal person, public authority, agency or other body which processes personal data **on behalf of** the controller.

Moreover, if two organisations jointly determine the purposes and means of the processing through an AI system, they may be considered as **joint controllers**. This may be the case, for example, where an organisation cooperates with another organisation in developing a product or service for which both parties provide personal data for the training and/or validation of the tool, and where they jointly determine the purpose of such processing and combine their technical resources, without one party processing personal data solely on the instructions of the other. The Table 1 below provides a schematic overview of different roles a data controller and data processor can have in an AI context.

Table 1 Different roles an organisation can play under the GDPR in an AI context

PHASE/ACTIVITY	WHO IS THE CONTROLLER?	WHO IS THE PROCESSOR?
DEVELOPMENT/ TRAINING/ VALIDATION	The person or organisation that develops, trains or validates the AI system and decides what personal data will be used to train the system (and therefore determines the purpose and means). If this organisation obtains a set of personal data from a third party, it will also have the status of controller when processing such data. If the development, training, validation or (further) development is outsourced to a third party organisation and this third party organisation decides which type of personal data is used in this regard, it becomes a controller.	The organisation to which the development, training, validation or (further) development is outsourced, provided that the client to whom such services are provided: (i) identifies the purpose of the processing activity; and (ii) determines the significant characteristics of the personal data to be processed. This is regardless of whether this client/controller transfers the personal data to the processor or the processor obtains it through its own channels and; (iii) the processor processes such data only for the purposes specified by the controller.

LAUNCH/ RELEASE/ **COMMISSIONING**	Any organisation that integrates an AI system into its product or service and thereby processes personal data for its own purposes. If the AI system (whether or not part of a wider product or service) is sold or licensed and already contains personal data, both organisations exchange personal data and are both controllers. Even if, for instance, a licensor makes a system available to a	Any organisation that makes an AI system available to a controller whereby the AI system is integrated into the latter's product or service, or any organisation that does so because it is necessary for the proper performance of its service, but that does not itself process personal data obtained from the controller for its own purposes. An organisation (service provider) that makes an AI system available to another organisation (user) is

	licensee and only the licensee is the controller (see on the right), the licensor still also becomes a controller when it processes personal data obtained from the licensee for its own purposes (e.g., to measure the efficiency of the AI system).	neither a processor nor a controller if: (i) this system is installed locally and stand-alone at the user's premises; and (ii) the service provider does not have access to the local installation, e.g., for maintenance.

What is the relationship between the GDPR and the AI?

The GDPR does not contain the term 'artificial intelligence', nor any terms expressing related concepts, such as intelligent systems, autonomous systems, automated reasoning and inference, machine learning or even big data. This does not, however, mean that the GDPR does not apply to training, testing, validation or deploying the AI systems. To the contrary, as presented in Figure 2, many provisions in the GDPR are very relevant to AI.

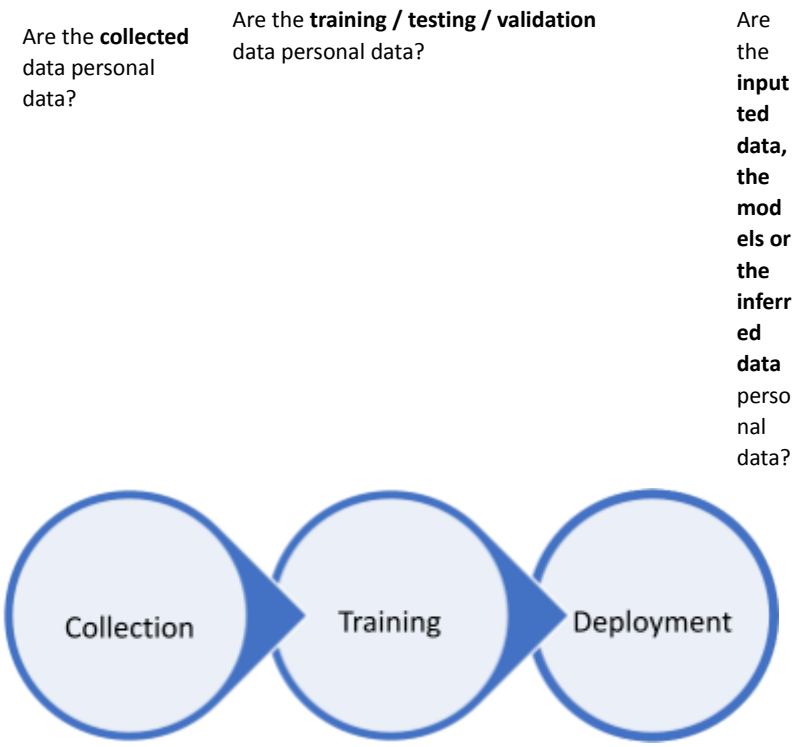


Figure 2 When does the GDPR apply to AI operations?

Generally speaking, many AI applications process personal data. On the one hand, personal data may contribute to the data sets used to train ML systems, namely, to build their algorithmic models. On the other hand, such models can be applied to personal data, to make inferences concerning particular individuals. Next, thanks to AI processing, personal data can be used to analyse, forecast and influence human behaviour. In the context of media, personal

data are often collected, processed and used for many purposes, among which automated personalisation of (recommendations for) content (e.g., news) and advertising (e.g., targeted advertisements).

What are the GDPR principles?

Lawfulness

Article 6(1) of the GDPR specifies that data processing is lawful only if it is based on one of six specified conditions set out in Article 6(1)(a) to (f) (see Figure 3 below).

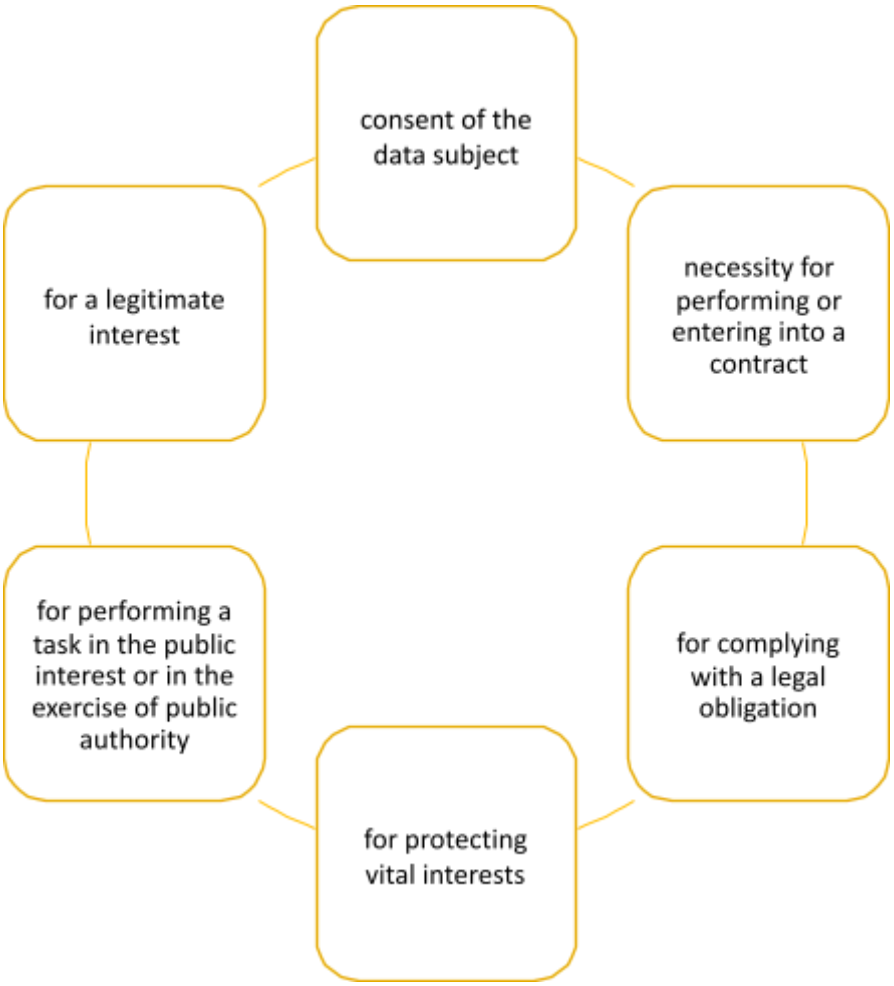


Figure 3 Lawful basis to process personal data

Identifying the appropriate lawful basis is of essential importance. Whenever you are processing personal data – whether to train a new AI system or make predictions using an existing one – you must have an appropriate lawful basis to do so.

Fairness

Two different concepts of fairness can be distinguished in the GDPR.[7] The first, which we may call 'information fairness' is strictly connected to the idea of transparency. It requires that data subjects are not deceived or misled concerning the processing of their data. Recital 71 points to a different dimension of fairness, i.e. what we may call 'substantive fairness', which concerns the fairness of the content of an automated inference or decision. Clifford and Ausloos notice

[7] Giovanni Sartor and others, *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study* (2020) <http://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf > accessed 15 June 2021.

that the notion of fairness can have two main meanings: fair balancing and procedural fairness.[8] Fair balancing is based on proportionality between data subjects' interests (e.g. the right to privacy, right to data protection) and necessity of purposes of the data controller. Procedural fairness refers to practical implementation of 'fairness' through specific procedures that can improve the level of transparency and lawfulness of a certain data processing in a specific context.[9]

Transparency

Articles 12, 13 and 14 of the GDPR contain the main general transparency obligations that controllers must comply with. There are also other, complementary transparency provisions such as 'data protection by design and by default' (Article 25), records of processing activities (Article 30) and data protection impact assessment (DPIA, Article 30).

In the event of **direct collection** of data from the data subject (Art. 13 of the GDPR), the controller must, at the time when personal data are obtained, provide the data subject with, inter alia, the following information:

- the identity and the contact details of the controller and, where applicable, of the controller's representative;
- the contact details of the data protection officer, where applicable;
- the purposes of the processing for which the personal data are intended as well as the legal basis for the processing;
- the recipients or categories of recipients of the personal data, if any;
- the period for which the personal data will be stored, or if that is not possible, the criteria used to determine that period;
- which rights the data subject has, including the rights a data subject has in the event of automated decision-making (i.e. right to human intervention, right to express an opinion and right to challenge the decision) and how these can be exercised;
- the right to lodge a complaint with a supervisory authority;
- the existence of automated decision-making, including profiling and meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

In case of **indirect** data collection (Art. 14 of the GDPR) e.g. through a third party, the same information must be communicated, along with:

- the categories of personal data concerned.
- from which source the personal data originate, and if applicable, whether it came from publicly accessible sources.

There are some exceptions to this information obligation in the case of indirect data collection as provided for in Article 14 of the GDPR. For example, this information does not have to be provided in cases where personal data are processed for scientific research purposes and (i) the

[8] Damian Clifford and Jef Ausloos, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130.

[9] Clifford and Ausloos (n 8).

provision of such information proves impossible or would involve a disproportionate effort, or (ii) the provision of such information is likely to render impossible or seriously impair the achievement of the objectives of that processing. Table 2 provides an overview of the information requirements specific to AI systems.

Table 2 Information requirements specific to AI systems

GDPR PROVISION	OBLIGATION
Art. 13(2)(f),	Obligation to inform about **the existence and use** of automated (individual) decision-making and profiling
Art. 14(2)(g),	Obligation to provide 'meaningful information on the **logic** involved'
Art. 15(1)(h)[10]	Obligation to inform about the '**significance** and the envisaged **consequences**' of this processing for the data subject
Art. 22 **Recital 71**	Obligation to provide **explanation** of the individual automated decision

Purpose limitation

Article 5(1)b GDPR stipulates that personal data shall be collected for specified, explicit and legitimate purposes and not further processed in a manner that is incompatible with those purposes. This means that the data should correspond to the aims justifying their collection, and the purpose must be sufficiently unambiguous and clearly expressed. Finally, they must be legitimate in the sense that they must match the legal expectations of data subjects (not to be confused with lawfulness).

Personal data collected for these purposes should not be 'further processed' in a manner which is incompatible with them. In principle, further processing, namely processing the data for another purpose than the purpose for which they were initially collected, should however be allowed where compatible with the initial purpose(s) (Recital 50 GDPR). Therefore, when personal data are further used for compatible purposes, 'no legal basis separate from that which allowed the collection of the personal data is required'. This is based on 'the reasonable expectations of data subjects based on their relationship with the controller as to the data's further use' (Recital 50 GDPR).

[10] Article 15(1)h is identical to Articles 13(2)f and 14(2)h: data subjects have a right to be informed about the existence of automated decision-making and to obtain meaningful information about the significance, envisaged consequences, and logic involved. However, there is a difference between the information requirements in Articles 13-14 and those in Articles 15. In the first case, the information must in principle be provided before or at the time of the processing of personal data. In the second case, the information will usually only be provided after the data subject requests it. The data subject can request this information at any time, including after the automated decision concerning him/her has been taken, with no deadline. As explained by the WP29 Guidelines *"Article 15 implies a more general form of oversight, rather than a right to an explanation of a particular decision."*

Data minimisation

Data minimisation is about asking whether the same purpose can be achieved with a narrower collection of data. Article 5(1)c GDPR requires to ensure that personal data are 'adequate, relevant and limited to what is necessary in relation to the purposes for which they are processed'.

Accuracy

Article 5(1)d of the GDPR requires controllers to ensure that the personal data are 'accurate and, where necessary, kept up to date; every reasonable step must be taken to ensure that personal data that are inaccurate, having regard to the purposes for which they are processed, are erased or rectified without delay'. Moreover, the Court of Justice of the European Union even pointed out that the accuracy requirements in light of privacy fundamental rights should ensure that the models and criteria used in the context of automated processing of personal data are 'specific and reliable' to fulfil the processing purposes.[11]

Storage limitation

Article 5(1)(e) GDPR contains the data storage limitation principle according to which personal data must not be kept for longer than is necessary for the legitimate purposes for which they are processed. A longer-than-necessary storage period is permitted only for processing for archiving achieving purposes in the public interest, scientific or historical research purposes or statistical purposes, subject to implementation of appropriate technical and organizational safeguards.

Integrity and confidentiality (security)

The integrity and confidentiality principles are essentially about information security. It requires controllers to ensure appropriate security when processing personal data and protect the data against unauthorized or unlawful processing, accidental loss, destruction or damage.

Accountability

The principle of accountability requires controllers and processors to be able to demonstrate that they have taken steps to comply with the obligations under the GDPR. WP29 Opinion on Accountability provides that accountability means showing how responsibility is exercised, demonstrated, and made verifiable.[12] In other words, responsibility needs to be demonstrated as working efficiently in practice to be able to develop sufficient trust.

3.2 The AI Act

The European Commission unveiled a proposal for a new Artificial Intelligence Act (AI Act) in April 2021 (see also Deliverable D2.1 for a detailed analysis). On 14 June 2023, Members of the

[11] Opinion of the Court (Grand Chamber) of 26 July 2017).

[12] WP29 Opinion 3/2010 on the principle of accountability.

European Parliament (MEPs) adopted their negotiating position on the AI Act.[13] The talks will now begin with EU Member States in the Council on the final form of the law. The aim is to reach an agreement by the end of this year; if adopted quickly, the AI Act will apply at the earliest in 2025.

Since the AI Act is still being negotiated, it is difficult to predict the final outcome. However, the key amendments proposed by the Parliament in relation to the Commission's version include:

- **An obligation to comply with "general principles"** which are applicable to all AI systems: all operators shall make their best efforts to develop and use AI systems or foundation models in accordance with the following principles: (i) human agency and oversight; (ii) technical robustness and safety; (iii) privacy and data governance. These principles steam from a high-level framework that promotes a human-centric European approach to ethical and trustworthy Artificial Intelligence (for details see deliverable D2.1).

- **Expanding prohibition on certain uses of AI systems** to include: facial recognition and other real-time remote biometric identification in publicly accessible spaces, predictive policing systems, biometric categorization using sensitive characteristics of natural persons, emotion recognition systems used in law enforcement, border management, workplace, and educational institutions and the creation of facial recognition databases on the basis of indiscriminate scraping of biometric data from social media or CCTV footage;

- **Imposing additional obligations for deployers of high-risk AI systems** (and not only on providers as proposed by the EC). The obligations require implementing human oversight, robustness and cybersecurity. New Article 29a requires to conduct a *fundamental rights impact assessment for high-risk AI systems* prior to putting a high-risk AI system into use. The assessment shall include, at a minimum, the following elements:

    - a clear outline of the intended purpose for which the system will be used;
    - a clear outline of the intended geographic and temporal scope of the system's use;
    - categories of natural persons and groups likely to be affected by the use of the system;
    - verification that the use of the system is compliant with relevant Union and national law on fundamental rights;

[13] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD))1.

  - the reasonably foreseeable impact on fundamental rights of putting the high-risk AI system into use;
  - specific risks of harm likely to impact marginalised persons or vulnerable groups;
  - the reasonably foreseeable adverse impact of the use of the system on the environment;
  - a detailed plan as to how the harms and the negative impact on fundamental rights identified will be mitigated.
  - the governance system the deployer will put in place, including human oversight, complaint-handling and redress.

- **Expanding the list of high-risk AI systems** is to include systems aimed at influencing voters in political campaigns or used in recommender systems of very large online platforms.

Additionally, in light of the recent developments related to the massive use of ChatGPT, the European Parliament proposes imposing **requirements on providers of foundation models** (see section 5.3 below).

The EP proposal has also introduced some provisions which guarantee **a greater personal data protection**. First, recital 45a provides that "the right to privacy and to protection of personal data must be guaranteed throughout the entire lifecycle of the AI system". This means that the principles of data minimization and data protection by design and by default, as set out in the GDPR "are essential when the processing of data involves significant risks to the fundamental rights of individuals". To this end, the amendment further provides that the "providers and users of AI systems should implement state-of-the-art technical and organizational measures in order to protect those rights. Such measures should include not only anonymization and encryption, but also the use of increasingly available technology that permits algorithms to be brought to the data and allows valuable insights to be derived without the transmission between parties or unnecessary copying of the raw or structured data themselves."

3.3 Data Governance Act (DGA)

The Data Governance Act (DGA) creates a framework for increased data availability and establishes conditions and frameworks for the re-use of data held by public sector bodies within the EU. The DGA also makes it easier for individuals and companies to make data voluntarily available for the common good, such as medical research projects (so-called data altruism).

The Act came into force on 23 June 2022, and it will be applicable from 24 September 2023. The DGA applies to "any digital representation of acts, facts or information", including personal data. Recital 4 DGA stipulates that in the event of a conflict between the DGA and Union law on the protection of personal data, such as GDPR, the latter should prevail. Despite this guarantee, the

DGA's interaction with the GDPR is problematic and is characterized by a number of legal questions.

As already provided in the deliverable D4.3, on 10 March 2021, the EDPB and EDPS adopted a joint opinion on the DGA proposal pointing out some of these inconsistencies. However, researchers point out that there are the divergence of definitions between the GDPR and the DGA ("data subject" vs "data holder), the uncertainty of the legal concepts used in the DGA, such as "data altruism" or "common interest", and the problematic relationship of consent.[14]

3.4 Digital Services Act (DSA)

The DSA has been published in the Official Journal as of 27 October 2022 and came into force on 16 November 2022.[15] The DSA will be directly applicable across the EU and will apply from 1 January 2024. From 17 February 2024, the DSA rules will apply for all regulated entities; by this date the EU Member States will have to establish Digital Services Coordinators.

The DSA is the result of a years-long drafting and negotiation process. It was intended as the EU's landmark piece of legislation for addressing illegal and harmful content and activity online. Overall, the main goal is to create a safer digital space in which the fundamental rights of all users of online intermediaries (such as online marketplaces, social networks, content-sharing platforms, app stores and search engines) are protected (see D6.2 "*Report on policy for content moderation*" for further information).

In the context of this Deliverable, the provisions regarding the access to data for researchers are particularly important. The DSA requires providers of Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs) to grant 'vetted researchers' access to data, subject to certain conditions. Data can be provided "for the sole purpose of conducting research that contributes to the detection, identification and understanding of systemic risks (…) and to the assessment of the adequacy, efficiency and impacts of the risk mitigation measures (…)"(art. 40(4)). Vetted researchers must meet certain criteria and procedural requirements in the application process. Importantly, they must be affiliated to a research organization or a not-for- profit body, organization or association (art. 40(12)). Many details around researchers' access to data through the DSA will be decided in delegated acts that have yet to be adopted by the European Commission.

The DSA has a potential to change the long-lasting problem of insufficient access to platforms' data for research communities. It is well-known that APIs conditions are often restrictive and its access can be shut down at any time with no explanation. That is the case of Twitter (now X). In February, the platform announced that it would close the free access to its API. Although the

[14] Lusine Vardanyan and Hovsep Kocharyan, 'The GDPR and the DGA Proposal: Are They in Controversial Relationship?' (2022) 9 European Studies 91.

[15] Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act), OJ L 277, 27.10.2022, p. 1–102.

change did not come into effect until the end of June, research on dis- and misinformation and social effects of the online platforms has become more problematic or even stopped entirely.

3.5 Data Act

While the Data Governance Act creates the processes and structures to facilitate the different types of data exchange, the Data Act complements the DGA and aims to clarify who can create value from data and under which conditions.[16] The EU policymakers reached a political agreement on the Data Act in June 2023.[17]

With this new legislation, the EC regulates data access, use and re-use to foster growth and innovation while preserving "European values" such as privacy, safety, security and "ethical standards". This legislation "meant to remove barriers to the circulation of non-personal data by regulating the rights and obligations of all economic actors involved in producing and consuming Internet of Things products"[18]. The text aims to solve the imbalance of powers between tech providers, users and smaller innovations actors.

So far, when EU consumers were buying connected devices, the question of data generated by these devices was not clear. The Data Act aims to "give both individuals and businesses more control over their data through a reinforced data portability right, copying or transferring data easily from across different services, where the data are generated through smart objects, machines and devices."[19] Moreover, the SMEs will now be protected against unfair contractual terms thanks to an unfairness test for contractual clauses.

The Data Act also has specific provisions in case of exceptional situations of high public interest (e.g. floods, wildfire). If the data from private companies appears necessary to address public emergency situation, it will be provided for free under certain conditions.

In terms of flexibility, the Data Act will enable more easily to move data and applications from one provider to another. Mandatory safeguards will be applicable to protect data held on cloud infrastructures in the EU.

The Data Act has strong links with the GDPR, especially the right to data portability enshrined in Art. 20 GDPR. The legislation "allows data subjects to move their data between controllers who offer competing services. (…) The Data Act will enhance this right for connected products so that consumers can access and port any data generated by the product, both personal and non-

---

[16] Charlotte Ducuing and Thomas Margoni, 'Data Act Blog Post Series: Introduction' (*CITIP blog*, 21 April 2022) <https://www.law.kuleuven.be/citip/blog/data-act-blog-post-series-introduction/> accessed 15 September 2023.

[17] Luca Bertuzzi, 'Data Act: EU Institutions Finalise Agreement on Industrial Data Law' (*www.euractiv.com*, 28 June 2023) <https://www.euractiv.com/section/data-privacy/news/data-act-eu-institutions-finalise-agreement-on-industrial-data-law/> accessed 15 September 2023.

[18] Bertuzzi (n 17).

[19] European Commission, 'Data Act – Questions and Answers*' (*Europa*) <https://ec.europa.eu/commission/presscorner/detail/en/qanda_22_1114> accessed 15 September 2023.

personal."[20] This is a real improvement, as the scope of this right under the GDPR was quite limited due to strict conditions contained in the GDPR (technical feasibility, consent or contractual lawful basis…).[21]

3.6 European Data Protection Supervisor (EDPS) and European Data Protection Board (EDPB)

The European Data Protection Board (EDPB) is an independent European body bringing together the national data protection authorities and the European Data Protection Supervisor (EDPS).[22] The EDPS is the European Union's (EU) independent data protection authority.[23] The two institutions provide guidance on how to interpret the GDPR rules including in an AI system context. They are also consulted regularly to provide opinion on new legislations.

Deliverable D4.3 has already described the concerns regarding the relationship between the AI Act and the GDPR as formulated in the EDPB-EDPS Joint Opinion 5/2021 on the proposal for an Artificial Intelligence Act. The opinion underlined that given that the development and use of AI systems will in many cases involve the processing of personal data, "ensuring clarity of the relationship of this Proposal to the existing EU legislation on data protection is of utmost importance".[24] The Opinion also emphasizes the need for a human-centric legal framework for AI in the EU.

In February 2022, the EDPB addressed a letter to the Justice Commissioner Didier Reynders in which it provides its opinion on the adaptation of liability rules to the digital age and artificial intelligence.[25] Overall, the document underscores the need for clear roles and responsibilities, explainable AI systems, security measures, and a focus on data protection principles in shaping liability rules for AI in the European context. It emphasises on the following elements:

1. **Clear definition of roles**: The EDPB highlights the need to clearly define the roles and responsibilities of AI system providers. This is crucial for aligning with existing regulations like GDPR and ensuring clarity between data controllers and processors.

2. **Explainability and Responsibility**: To address the challenge of assigning responsibility in AI cases, the EDPB emphasizes the need to design the AI systems with built-in

[20] European Commission (n 19).

[21] Teodora Lalova-Spinks and Daniela Spajić, 'The Broadening of the Right to Data Portability for Internet-of-Things Products in the Data Act: Who Does the Act Actually Empower? (Part I)' (*CITIP blog*, 16 June 2022) <https://www.law.kuleuven.be/citip/blog/the-broadening-of-the-right-to-data-portability-for-internet-of-things-products-in-the-data-act-part-i/> accessed 15 September 2023.

[22] 'EDPB | European Data Protection Board' <https://edpb.europa.eu/edpb_en> accessed 15 September 2023.

[23] 'About | European Data Protection Supervisor' (23 August 2023) <https://edps.europa.eu/about-edps_en> accessed 15 September 2023.

[24] EDPB-EDPS Joint Opinion 5/2021 on the proposal for an Artificial Intelligence Act.

[25] EDPB, 'EDPB Letter to the European Commission on Adapting Liability Rules to the Digital Age and Artificial Intelligence (AI)' (22 February 2022) <https://edpb.europa.eu/our-work-tools/our-documents/letters/edpb-letter-european-commission-adapting-liability-rules_en>.

explainability. This includes human supervision, transparency, and the ability to explain AI results and potential claims.

3. **Security Measures**: Providers of AI systems should be responsible for equipping users with tools to mitigate known and new types of attacks. Security should be integrated throughout the AI system's lifecycle, even if it doesn't directly handle personal data.

4. **Documentation and Remedies**: AI systems should include accessible documentation to help data controllers understand system failures, especially those leading to data breaches. The EDPB calls for legal remedies for affected individuals following system failures or successful attacks.

5. **Data Protection Principles**: The EDPB stresses the importance of applying data protection principles effectively in AI usage. Inaccuracies and unfair algorithmic decisions can harm individuals. They recommend assessing data quality and fairness in AI decision-making and emphasize measurability to build trust and reduce negative impacts from erroneous decisions.

The EDPS seems extremely active on the topic as it recently released another interesting update on AI. For instance, in June 2023, the EDPS published a blogpost in which the institution highlights the significance of explainable AI (XAI) in making AI more transparent and accountable, while also addressing the challenges and risks associated with its implementation.[26] Collaboration among experts and compliance with data protection regulations are crucial for the responsible development and use of AI. The blog reminds that AI should be designed to serve humanity and align with human values (rec. 4 GDPR).

In July 2023, the EDPS set up a task force on Artificial Intelligence to ensure that this technology is used in full respect of data protection law.[27]

[26] 'Explainable Artificial Intelligence Needs Human Intelligence | European Data Protection Supervisor' <https://edps.europa.eu/press-publications/press-news/blog/explainable-artificial-intelligence-needs-human-intelligence_en> accessed 21 September 2023.

[27] European Data Protection Supervisor, 'Reshaping the EDPS to Tackle Data Protection Challenges' (23 August 2023) <https://edps.europa.eu/press-publications/press-news/press-releases/2023/reshaping-edps-tackle-data-protection-challenges> accessed 15 September 2023.

4 EU legal and ethical framework for the use of AI applications in media

4.1 Introduction

Data protection framework

As mentioned in the Deliverable *D4.3 - Initial analysis of the legal and ethical framework of trusted AI,* the GDPR does not contain the term 'artificial intelligence', nor any terms expressing related concepts, such as intelligent systems, autonomous systems, machine learning or even big data. This does not, however, mean that the GDPR does not apply to training, testing, validation or deploying the AI systems, including in the media sector. To the contrary, many provisions in the GDPR are very relevant to AI. In the context of media, personal data are often collected, processed and used for many purposes, among which automated personalisation of (recommendations for) content (e.g., news) and advertising (e.g., targeted advertisements). In what follows, we present three practical use-cases illustrating personal data considerations and ethical dilemmas while using personal data in AI media applications.

Ethical framework

It is worth recalling a recent research on AI ethics which presents a more critical stance towards a principle-based approach to AI ethics. In a thought-provoking article, Munn criticizes "the flood" of AI guidelines and codes of ethics for containing "meaningless principles which are contested or incoherent, making them difficult to apply; (...) isolated principles situated in an industry and education system which largely ignores ethics; and (...) toothless principles which lack consequences and adhere to corporate agendas".[28] The statement has been discussed by the experts during the Theme Development workshop on "Trusted AI - The future of creating ethical and responsible AI systems", [29] which was co-organized by AI4Media on 13 of September 2023. The workshop participants recognized the following key challenges when it comes to ethical AI:

- Proliferation of AI ethics guidelines, a lack of actual impact the AI ethics guidelines have, a lack of enforcement mechanisms in case of non-compliance, and no robust regulatory mechanism to govern ethical AI.
- Imbalance of funding between private sector and public sector and a reactive (instead of anticipatory) approach to a more anticipatory approach to technological developments;

[28] Munn, L. The uselessness of AI ethics. *AI Ethics* 3, 869–877 (2023). https://doi.org/10.1007/s43681-022- 00209-w

[29] https://www.vision4ai.eu/tdw-trusted-ai/

- Limits of principle-based approaches to AI Ethics and a lack of clarity on how to evaluate and balance values against each other, how to implement them in technical systems, and how to enforce them in practice. [30]

The discussants argued that in order to overcome these challenges, a new set of interdisciplinary skills and on-going governance is required to embed ethics in the entire cycle of AI development: from concept development to evaluation. Responsible development of technology requires a meaningful involvement of affected stakeholders from the phase of question articulation/ problem definition. There is a clear need for an embedded-ethics approach which incorporates reflections on potential consequences of AI development throughout the whole design and development process. The participants also pointed out to value-sensitive design-like approaches such as Values that Matter approach from the University of Twente.[31] These echo the recommendations by Bender et al. which argue that "value sensitive design provides a range of methodologies for identifying stakeholders (…) to identify their values, and designing systems that support those values. These include such techniques as envisioning cards, the development of value scenarios, and working with panels of experiential experts."[32] In order to mitigate the risks that come with the creation of increasingly large language models, researchers should therefore shift to a mindset of careful planning, along many dimensions, before starting to build
either datasets or systems trained on datasets.[33] It was also suggested to re-focus the conversation from high-level principles to AI justice. AI justice is a concept which "reframes much of the discussion around 'AI ethics' by drawing attention to the fact that the moral properties of algorithms are not internal to the models themselves but rather a product of the social systems within which they are deployed." [34]

4.2 Fact-checking and data protection

A use-case

[30] The 2nd cross-cutting Theme Development Workshop (TDW) on "Trusted AI: The Future of Creating Ethical and Responsible AI Systems", jointly organised by AI4Media, ELISE, ELSA, euRobin, HumanE-AI-Net, CLAIRE, TAILOR and VISION, took place on 13 September 2023, with the aim of developing and identifying the most promising and emerging themes related to the overarching concept of Trustworthy AI. The breakout session 8 "AI Ethics: from principles to practice" was moderated by Lidia Dutkiewicz from KU Leuven, with two experts: Isabela Rosal Santos from KU Leuven and Imre Bard from the Dutch National Laboratory on AI in Education. The full report on the key findings from the workshop will be made available at https://www.vision4ai.eu/tdw-trusted-ai/.

[31] Merlijn Smits and others, 'Values That Matter: Mediation Theory and Design for Values', *Research Perspectives in the era of transformations: Conference proceedings* (Academy for Design Innovation Management 2019) <https://research.utwente.nl/en/publications/values-that-matter-mediation-theory-and-design-for-values> accessed 21 September 2023.

[32] Emily M Bender and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 27 September 2023.

[33] Bender and others (n 32).

[34] Iason Gabriel, 'Toward a Theory of Justice for Artificial Intelligence' (2022) 151 Daedalus 218.

Sarah works as a fact-checker in a fact-checking NGO. She was conducting a study on the causes of the hyperactivity of certain users on Twitter (now X) to analyse whether they were linked with disinformation campaign and foreign interference. To do so, she uses machine learning (ML) to analyse the personal data of 55,000 Twitter users, and to categorise more than 3,300 accounts according to their political affiliation or their membership of political movements. To be transparent about the study and its methodology, the NGO decided to publish Excel files containing the raw personal data of the Twitter accounts analysed.

This use case has been inspired by the EU Disinfolab case (Decision of the Belgian data protection authority 13/2022 of 27 January 2022, DOS-2018-04433)

Figure 4 Fact-checking and data protection use case

Data protection considerations

Reuse of data available on social media

The fact that personal data is publicly available on social networks does not mean that they can be freely re-used irrespective of the GDPR. It is useful to see 'public' Twitter data as private data on public display which can be (re)used if there is a lawful basis for processing. Broadly speaking, these are: consent, contract, legal obligation, protection of vital interests, public interest, or a legitimate interest (see D4.3 for more detail).

For the collection of data, Art. 12-13 of the GDPR specify in more detail when that information needs to be communicated to a data subject. In case of indirect data collection (Art. 14 of the GDPR), e.g., through a third party (such as through outsourced software), the same information must be communicated, along with the categories of personal data concerned and an information from which source the personal data originate.

There are some exceptions to this information obligation in the case of indirect data collection as provided for in Article 14 of the GDPR. For example, this information does not have to be provided in cases where personal data are processed for scientific research purposes. Sarah could arguably rely on a journalistic/scientific research exception and therefore does not fall under the obligation to inform individuals individually about the personal data processed for the study. This exception would only apply provided that appropriate safeguards (technical and organizational) are met:

- internal documentation (processing log, prior impact assessment);
- external documentation (privacy policy) on the data processing methodology;
- security and confidentiality obligations: anonymization/pseudonymization.

If these conditions are not met, Sarah could not claim to benefit from the scientific processing exception.

Other elements which would need to be considered are:

(i) whether the data processing constitutes a 'high risk' operation

The following list (Figure 5) provides a few examples which require to complete a data protection impact assessment (DPIA) as they are 'likely to result in high risk'. These should not be taken as definitive or exhaustive. Since the data of 55,000 users was processed, this is likely to constitute a 'large-scale' processing.

(ii) whether sensitive data are processed

It is worth mentioning, that as provided by the UCL Guidelines, "since the researcher sampling the API stream of Tweets has little control over the content of the received Tweets at the point of reception, the data should be treated as if high-sensitivity/special- category at the point of ingestion, and appropriate protocols and data protection management put in place to address this" (UCL). In this regard, political profiling of the Twitter users concerned falls under a category of processing of sensitive data.

Type of processing operation(s) requiring a DPIA	Description	Non-exhaustive examples of existing areas of application
Innovative technology	Processing involving the use of new technologies, or the novel application of existing technologies (including AI). A DPIA is required for any intended processing operation(s) involving innovative use of technologies (or applying new technological and/or organisational solutions) when combined with any other criterion from WP248rev01.	• Artificial intelligence, machine learning and deep learning • Connected and autonomous vehicles • Intelligent transport systems • Smart technologies (including wearables) • Market research involving neuro-measurement (i.e. emotional response analysis and brain activity) • Some IoT applications, depending on the specific circumstances of the processing
Large-scale profiling	Any profiling of individuals on a large scale	• Data processed by Smart Meters or IoT applications • Hardware/software offering fitness/lifestyle monitoring • Social-media networks • Application of AI to existing process

Invisible processing	Processing of personal data that has not been obtained direct from the data subject in circumstances where the controller considers that compliance with Article 14 would prove impossible or involve disproportionate effort (as provided by Article 14.5(b). A DPIA is required for any intended processing operation(s) involving where the controller is relying on Article 14.5(b) when combined with any other criterion from WP248rev01	• List brokering • Direct marketing • Online tracking by third parties • Online advertising • Data aggregation/data aggregation platforms • Re-use of publicly available data

Figure 5 Examples of processing 'likely to result in high risk'. Source: ICO

For these reasons, to benefit from the research exemption, appropriate safeguards for the rights and the persons concerned should be adopted. In this case, it would have included a register of processing activities (because of sensitive data), external privacy policy on the methodology and confidentiality measures.

Publication of raw data from Twitter accounts

First off, researchers need to consider Twitter's guidelines and conditions for re-using the Tweets. Twitter permits the sharing of Tweet IDs and User IDs in a dataset for others to use (for academic research the number is currently unlimited). A limited number of 'hydrated' Tweets can be shared, but only privately, and the dataset creator who is sharing their dataset must ensure that the recipient has agreed to the Twitter terms before doing so. Twitter places particular restrictions on the form in which Tweets may be published, requiring certain items of data to be retained in the published form.[35]

From a GDPR perspective, a separate lawful basis for personal data processing must be established for the publication of the data. Article 6(1)(f) legal basis is the necessity of the data processing for the purposes of the legitimate interests, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject. In short, there are three elements to the legitimate interest lawful basis which a data controller has to comply with. To answer the question whether Sarah has a legitimate interest in processing personal data, she should pass three tests: i) the 'purpose test', namely identifying a legitimate

[35] Nicolas Gold, 'Using Twitter Data in Research Guidance for Researchers and Ethics Reviewers', available at: https://www.ucl.ac.uk/data-protection/sites/data-protection/files/using-twitter-research-v1.0.pdf.

interest; ii) the 'necessity test', showing that the processing is necessary to achieve it; (iii) the 'balancing test', balancing the necessity against the individuals' interests, rights and freedoms.

It could be argued that Sarah's research, as a journalist/researcher working in the disinformation domain, serves a legitimate interest purpose. Then, as regards the criteria of necessity, the purpose of transparency about research results and methodology, could perhaps be achieved in a way that is less harmful to the interests, rights and freedoms of the persons concerned. Such less harmful way could consider publishing duly pseudonymized data and by providing for access restrictions (e.g. on a case-by-case basis upon request). Another option would be to limit access to the data to accredited journalists to achieve the purpose, namely, to justify the methodology of the research. As regards the balance of interest, it is necessary to balance the contribution to a debate of general interest, and the reputation of the persons concerned. The people concerned could not reasonably expect that their sensitive data (such as political affiliation) will be publicly published, despite making them 'publicly' available on Twitter. The security measures are therefore of key importance here: pseudonymizing the data or limiting the access to the files.

Therefore, notwithstanding the moral legitimacy of the purposes pursued by the researcher, releasing raw and un-pseudonymised personal data containing sensitive personal data can be considered as a disproportionate means to demonstrate the research methodology.

Ethical considerations of using data

The dynamic character of Twitter's datasets

The Twitter contents change regularly: there are new Tweets, new responses, but also the deletion and other user-driven changes to the status of available information. Twitter expects those who use its information to respect the changes that users make. The UCL Guidelines point out that this means that the compliance with ethical and privacy requirements is not a one-off exercise, but it must be applied at every use and regularly during retention, not just at the outset of a research study. This has implications for research design.[36] In addition, Twitter's synchronisation requirement means that should a user delete or protect a Tweet that has been quoted in a paper, that paper would need to be modified to remove it.[37] This is a well-recognised issue in the ethics literature.[38]

[36] Ibid.

[37] Ibid.

[38] see Matthew L Williams, Pete Burnap and Luke Sloan, 'Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation' (2017) 51 Sociology 1149.

Consent and Expectations

Another ethical consideration is the need to identify and respect the stakeholders in the research. Although Twitter makes clear to its Tweeters, at the point of signing up to use its services, that their data may be used for research, the Developer API agreement provides that users retain control over the public disposition of their data, and the autonomy to change the status of their Tweets. According to the Menlo Report, the key primary ethical principle in this context is autonomy: the right of participants to be treated as independent moral agents and to determine their own best interests.[39] To this end, retention and use of data which is no longer on the public display breaches users' autonomy since it does not respect their current wishes to withdraw data from availability. Moreover, a research by Williams et al. shows that while there is a general lack of concern from users over their posts being used for research purposes (with university research attracting least concern), 80 per cent of respondents expected to be asked for their consent ahead of their Twitter content being published, and over 90 per cent stated they expected anonymity in publication.[40] In addition, one needs to consider the situation of the persons whose data is disclosed implicitly by a Tweeter. In many cases, there is no evidence that such a person has consented to their information being publicly available on Twitter and there re-used for research purposes.

4.3 Recommender systems and data protection

A use case

Tom works for news broadcasting organization which is a public service media. He is part of a team developing the personalized recommender system on the news organization website. The system filters, suggests, and prioritises content based on previous or similar users' behaviour, explicitly stated user preferences, popularity metrics, and other content-specific features.

Figure 6 Recommender systems and data protection use case

Data protection considerations

Privacy risks

[39] Erin Kenneally, David Dittrich 'The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research by Erin Kenneally, David Dittrich :: SSRN' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2445102> accessed 21 August 2023.
[40] Williams, Burnap and Sloan (n 38).

Recommender systems can be explicit or implicit also known as self-selected personalisation or pre-selected personalization.[41] Users can specifically provide information. In this way, users build their own profile specifying their likes and dislikes, or provide information such as age and gender about themselves. As a way of example, BBC iPlayer requires user sign-in to offer recommendations (Figure 7), requiring data such as user's email, date of birth and location.

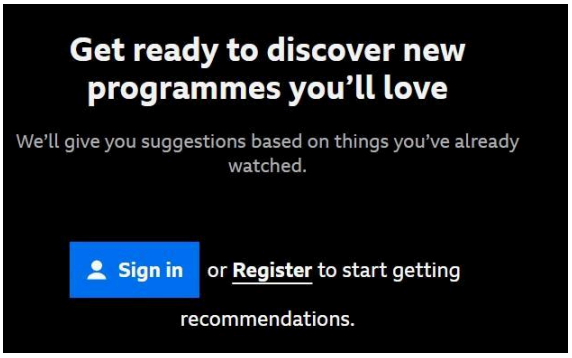


Figure 7 BBC iPlayer

The information supplied can also be automatically collected as the result of users interacting with the recommender systems and making choices based on recommendations. Again, BBC iPlayer, for instance, recommends programmes that the organization thinks the user will like. These recommendations are based on things already watched. It also recommends things that people with similar tastes are watching. It should be noted the data that is collected are considered personal data if the related person can "reasonably likely" be identified. Hildebrandt points out that due to the linkability of much behavioral data with other data this will often be the case, triggering the applicability of the GDPR.[42]

Based on the information provided in a use case, it can be assumed that both explicit and implicit data collection takes place.

The data gathering and operation of recommendation systems can pose **direct and indirect privacy risks**. Direct privacy risks steam from non-compliance with existing privacy regulations and/or malicious use of personal data. Indirect risks result from data leaks, deanonymisation of datasets or unwanted exposure of inferred sensitive characteristics to third parties.[43]

[41] see Frederik J Zuiderveen Borgesius and others, 'Should We Worry about Filter Bubbles?' (2016) 5 Internet Policy Review <https://policyreview.info/node/401> accessed 21 September 2021.).

[42] Mireille Hildebrandt, 'The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems' (2022) 5 Frontiers in Artificial Intelligence <https://www.frontiersin.org/articles/10.3389/frai.2022.789076> accessed 18 August 2023..

[43] Silvia Milano, Mariarosaria Taddeo and Luciano Floridi, 'Recommender Systems and Their Ethical Challenges' (2020) 35 AI & SOCIETY 957.

Friedman et al. classified privacy risks in recommender systems as follows[44] (Figure 8):

Adversary	*Direct access to existing data*	*Inference of new data*
Recommender system	Unsolicited data collection Sharing data with third parties Unsolicited access by employees	Exposure of sensitive information Targeted advertising Discrimination
Other users	Leaks through shared device or service	Inference from the recommender output
External entities	Lawful data disclosure Hacking Re-identification of anonymized data	Exposure of sensitive information

Figure 8 Privacy risks in recommender system

Information obligations

As regards data protection rights and obligations, Articles 13-14 GDPR provide a detailed list of what information (explanation) should be provided to users. Data subjects should be informed in detail what aspects of their experience are personalised and how. Article 15 complements the information obligations, by granting data subjects an explicit right to obtain additional, *ex post*, information. This so-called "right of access" includes an information on the purposes of the processing, the categories of personal data concerned, the recipients or categories of recipient to whom the personal data have been or will be disclosed, the envisaged period for which the personal data will be stored, and the right to lodge a complaint and a right to erasure. Moreover, in case of automated decision-making, including profiling, a meaningful information about the logic involved, as well as the significance and the envisaged consequences of such processing for the data subject.

In other words, data controllers, and in the case in hand, the developers of the recommender systems, should be transparent about the full set of personal data they use to create the profile, where they got each piece of information from, what the exact purposes of the profiling are, and who it has been (or might be) shared with.

In case of recommender systems, it is likely that such a processing constitutes an "automated individual decision-making" (Art. 22(1) GDPR). The analysis in the ATAP project suggests that firstly, it is important to establish whether profiling is in fact taking place.[45] To determine this, one has to ask: does the news recommender gather information about (groups of) individuals and evaluates their characteristics or behaviour patterns in order to place them into a certain category or group, in particular to analyse and/or make predictions about, for example, their

[44] Arik Friedman and others, 'Privacy Aspects of Recommender Systems' in Francesco Ricci, Lior Rokach and Bracha Shapira (eds), *Recommender Systems Handbook* (Springer US 2015) <https://link.springer.com/10.1007/978-1-4899-7637-6_19> accessed 18 August 2023.

[45] Pierre Dewitte et al., 'Interdisciplinary Problem Formulation - ATAP - Work Package 1/Deliverable 1'.

interests or likely behaviour? If so, to what extent do the recommendations produce legal effects or similarly significantly affect data subjects?[46]

To implement these requirements in practice, users affected by the recommendations should first receive a high-level indication of how and why they are seeing specific news content, and second, they should have the opportunity to see more details for each specific post.

Users should also receive explanations. In the case at hand, one way of providing such explanations about how recommender system work, is experimental design or a concepts of "counterfactual explanations": "unlike existing approaches that try to provide insight into the internal logic of black box algorithms, counterfactual explanations do not attempt to clarify how decisions are made internally. Instead, they provide insight into which external facts could be different in order to arrive at a desired outcome". [47]

Data protection by design and by default

Data protection by design and by default (art. 25 GDPR) obligations may require news recommenders to develop clear visualisations and controls over how content is in fact arranged and presented to the individual. A useful way for facilitating data subjects' understanding on how recommender systems work, are 'privacy dashboards', which have a similar function to model cards.[48]

As a side note, it is worth mentioning that recently adopted DSA regulation has a specific set of obligations for the providers of online platforms which go beyond data protection considerations (see deliverable D6.2 "*Report on policy on content moderation*").

Ethical considerations

Milano et al. provide a taxonomy of ethical considerations in recommender systems as well as possible solutions (Table 3). These include: (i) privacy violations in a form of unfair or otherwise malicious uses of personal data to target individual users; (ii) issues of personal autonomy; (iii) nudging; (iv) the opacity about which and whose values are at stake in recommender systems.[49]

[46] Ibid.

[47] Sandra Wachter, Brent Mittelstadt and Chris Russell, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' <http://arxiv.org/abs/1711.00399> accessed 21 August 2023. See also D4.3 and Graziani, M., Dutkiewicz, L., Calvaresi, D. et al. 'A global taxonomy of interpretable AI: unifying the terminology for the technical and social sciences', Artif Intell Rev 56, 3473–3504 (2023). https://doi.org/10.1007/s10462-022-10256-8.

[48] 'Model Cards - OECD.AI' <https://oecd.ai/fr/catalogue/tools/model-cards> accessed 21 August 2023.

[49] Milano, Taddeo and Floridi (n 43).

Table 3 Ethical issues of recommender systems and possible solutions. Adopted from Milano, Taddeo, and Floridi 2020.

Area of concern	Issues	Solutions
Content	Recommendation of inappropriate content	User-specified filters
		Demographic or geographical filters
Privacy	Unauthorised data collection and storage	Data storage in separate databases
	Data leaks	Anonymization, encryption
	Unauthorised interferences	Legislation (GDPR)
Autonomy and personal identity	Behavioural traps	Increase the transparency of user categorisation
	Encroachment on sense of personal identity	
Opacity	Black-box algorithms	Factual explanations
	Uninformative explanations	
	Feedback effects	
Fairness	Observation bias	Multi-stakeholder recommendations framework
	Population imbalance	
Social effects	Lack of exposure to contrasting viewpoints	Recommender personae
	Feedback effects	Serendipitous recommendations

4.4 Facial recognition and data protection

A use case

The public broadcaster AAB, a renowned media organization, is exploring the development of a facial recognition tool to enhance the quality and efficiency of its news programs archiving. With an extensive archive of news content and a constant need to identify celebrities appearing in various news segments, this tool aims to automate and streamline the process of recognizing and tagging celebrities in their news programs.

Figure 9 Facial recognition and data protection use case

Data protection considerations

Variety of functions

First off, it should be noticed that facial recognition can fulfil two distinct functions: (i) the **authentication** of a person, aimed at verifying that a person is who she or he claims to be by comparing of two templates; (ii) the **identification** of a person, aimed at finding a person among a group of individuals, in a specific area, an image or a database by comparing one template with a database of templates or samples.[50]

Facial recognition use grows exponentially in different sectors. For instance, to unlock smartphone or to tag people on social media platforms (facial recognition authentication) but also in the public spaces or in law enforcement (facial recognition identification). In relation to the media sector, facial recognition could be helpful for content management and organisation, content moderation or video editing or copyright protection screening but it could also be used for unethical and legally problematic uses. For instance, using facial recognition to analyse viewers' preference, habits, attention and mood about a certain content. For instance, in 2015, Usher released an exclusive music video on Tidal; during the viewing of the video, users' webcam were turned on. If they would turn their head away, the message "Don't Look Away" would pop up and the music video would be paused.[51] Such applications raise significant concerns regarding privacy and data protection. Often, users agree to the terms and conditions of platforms without carefully reading or fully understanding the extent of their consent, resulting in unexpected or unwanted outcomes.

Privacy concerns

Facial recognition technology is not explicitly regulated yet and there are doubts about its data protection and ethical compliance.[52] However, it does not mean that existing legal regimes, including the GDPR and the EU Charter of fundamental rights and the primary law, are not already applicable to this technology. Therefore, the facial recognition technologies in the EU must comply with the GDPR to be legally allowed.

Facial recognition technology processes biometric data considered as sensitive data by the GDPR. The biometric data is defined as data 'relating to the physical, physiological or behavioural characteristics of a natural person, which allow or confirm the unique identification of that natural person'.[53] The processing of biometric data is subject to strict limitation since processing of sensitive personal data is generally prohibited unless one of the exemptions of Art. 9(2) of the GDPR applies. While not all images containing faces are personal data, they become personal

[50] EDPB, Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement.

[51] Josh Hymowitz, 'Facial Recognition and Mood Detection in Media' (Medium and Media-Nxt: The Future of Media, 2 November 2020) <https://medium.com/media-nxt/facial-recognition-and-mood-detection-in-media-how-the-trend-is-emerging-in-advertising-marketing-d223c4a2d8e2> accessed 28 September 2023.

[52] Wojciech Wiewiórowski, 'Facial Recognition: A Solution in Search of a Problem?' (*European Data Protection Supervisor*, 28 October 2019) <https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem> accessed 21 September 2023.

[53] Art. 2(14) GDPR.

data when they clearly show the faces to allow the identification of individuals. To clearly enable identification, several parameters enter into account: light, pose, quality of the image.[54] The problem with face images is that "unlike other biometric data, face recognition data can be collected without physical intrusion, and even without the individual's knowledge, solving issues related to bodily integrity, but creating additional problems on the matter of informed consent and privacy"[55].

It is certain that this technology presents some benefits for instance in terms of efficiency and security however, the privacy and ethical challenges cannot be overlooked.[56] The Civil Liberties Union for Europe (Liberties), which is a watchdog that safeguards human rights in the EU, provided a list of seven biggest privacy concerns in relation to facial recognition (see Table 4).[57]

Table 4 Liberties list of privacy concerns.

Facial Recognition Privacy Risks
1. *Improper data storage*
2. *Misuse of data*
3. *Infringement on individual privacy*
4. *Infringement on freedom of speech and association*
5. *Lack of transparency*
6. *Normalisation of facial recognition*
7. *Wide Accessibility of facial recognition*

These concerns will be further elaborated in the below sections.

Improper data storage

Facial images are easily collected in public places, but the problem lies in the insecurity of biometric databases.[58] If these databases get hacked, your personal information can be used for identity theft or harassment. Biometric data, like facial features, cannot be changed if compromised, making it a serious concern, as highlighted by the Clearview AI controversy. In

[54] Catherine Jasserand, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part I)' (*CITIP blog*, 28 April 2022) <https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-i/> accessed 21 September 2023.
[55] Vera Lúcia Raposo, '(Do Not) Remember My Face: Uses of Facial Recognition Technology in Light of the General Data Protection Regulation' (2023) 32 Information & Communications Technology Law 45.
[56] Raposo (n 55).
[57] Liberties.EU, '7 Biggest Privacy Concerns Around Facial Recognition Technology| LibertiesEU' (*Liberties.eu*, 25 October 2022) <https://www.liberties.eu/en/stories/facial-recognition-privacy-concerns/44518> accessed 20 September 2023.
[58] Liberties.EU (n 57).

2020, the New York Times revealed that the US facial recognition company Clearview sold three billion face images to police authorities and private companies.[59] Clearview collected the images but also associated information (e.g. the source of the image (URL), the geo-localisation, and sometimes the name of the individuals).[60]

V.L. Raposo also acknowledges the need for security measures. However, she nuances this need by putting forward that facial recognition technologies are creating templates from photos and that in case of a data breach due to current technology limitations, reverse engineering of these codes to obtain a picture is extremely challenging, if not impossible.[61] The security measures necessary to safeguard the integrity of this database include but non-exclusively the following: data protection impact assessment, photos and templates which should be stored separated, and a need to have decentralized model and encryption of the data in the centralized model.[62]

Bias and errors

Facial recognition technology exhibits bias, particularly favouring white men over other groups, since face recognition models are mainly trained with such data.[63] Bias in facial recognition can lead to false results and errors leading to problematic consequences such as arrests or incorrect convictions or other forms of discrimination. For instance, a report from video surveillance researchers at IPVM[64] warned that Huawei worked on a facial recognition system to monitor and track China's Uighur minority.[65]

Lack of lawful basis

Facial recognition technology's current use is invasive and intrusive by collecting data without individuals' consent or awareness, such as being filmed in public spaces. While face images to train the facial recognition models and systems can be freely accessible on the internet and easily collected thanks to scrapping techniques, however they are not freely re-usable according to the GDPR and a lawful basis is necessary to allow the processing (Art. 6 GDPR).

---

[59] Kashmir Hill, 'The Secretive Company That Might End Privacy as We Know It' *The New York Times* (18 January 2020) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> accessed 21 September 2023.

[60] Jasserand (n 54).

[61] Raposo (n 55).

[62] Raposo (n 55).

[63] Liberties.EU (n 57).

[64] ipvideomarket, 'Huawei / Megvii Uyghur Alarms' (*IPVM*, 8 December 2020) <https://ipvm.com/reports/huawei-megvii-uygur> accessed 20 September 2023.

[65] Russell Brandom, 'Huawei Worked on Facial Recognition System to Surveil Uighurs, New Report Claims' (*The Verge*, 8 December 2020) <https://www.theverge.com/2020/12/8/22163499/huawei-uighur-

surveillance-facial-recognition-megvii-uyghur> accessed 20 September 2023.

Consent cannot be used as a legal basis when individuals are not even aware of the collection of their personal data. Indeed, "consent must be expressed through positive agreement, not one inferred from the data subject's conduct".[66]

The legitimate interests require a careful balancing exercise between the controller's and data subject's interests. In the media use case context, it is fair to note that commercial interests are included within the scope of the legitimate interest, as confirmed by the WP29 opinion.[67] The list includes the exercise of the right to freedom of expression or information, including in the media and the arts. However, when it comes to biometric data a lawful basis for commercial purposes is more difficult to find. This is also because legitimate interest is not included in the exception to general prohibition of processing sensitive personal data in Art. 9(2) of the GDPR. Art. 9(2)e of the GDPR states that when the processing relates to personal data which are manifestly made public by the data subject, the processing of sensitive data can be accepted. However, the EDPB in its guidelines on the processing of personal data through video devices states that: "the mere fact of entering into the range of the camera does not imply that the data subject intends to make public special categories of data relating to him or her".[68] This exception has to be interpreted narrowly. In addition, the data subjects must reasonably expect the further use of their personal data being publicly accessible, and "it's hard to argue that individuals who have published their data on different social media platforms could reasonably expect that their data would be further processed for biometric identification purposes and stored in a database made available to users such as law enforcement authorities".[69] In addition, the EDPB added that determining the source of images shared on social media platforms, whether they were posted by a third party (like a friend or acquaintance) or by the individuals depicted in the images, can be a challenging task.[70]

The EDPS has also raised concerns about whether there could be a valid legal basis for the application of such technology given that it relies on the large-scale processing of sensitive data.[71] The French DPA has also questioned the possibility of using the legitimate interest basis for personal data processing that is massive, especially when it involves obtaining from the Internet information concerning millions of people. It found that: "individuals who have published photographs of themselves on websites, or consent to such publication with another

[66] Raposo (n 55).
[67] Raposo (n 55).
[68] European Data Protection Board, Guidelines 3/2019 on processing of personal data through video devices 2020.
[69] Catherine Jasserand, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part II)' (*CITIP blog*, 5 May 2022) <https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/> accessed 21 September 2023.
[70] European Data Protection Board, 'Guidelines 8/2020 on the Targeting of Social Media Users | European Data Protection Board' (13 April 2021) <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-82020-targeting-social-media-users_en> accessed 21 September 2023.
[71] Liberties.EU (n 57).

data controller, do not expect that they will be reused for the purposes pursued by the company, i.e., the creation of facial recognition software (which combines the image of an individual with a profile containing all the photographs in which they appear, the information those photographs contain as well as the websites on which they are located) and the marketing of this software to law enforcement authorities."[72]

Lack of transparency

So far, the use of this technology has been unclear. As pointed out by the EDPS, there is a lack of clarity regarding how data collectors utilize the information, who possesses the ability to access it, where it gets sent, the duration of its storage, the methods employed to construct profiles, and the individuals responsible for automated decision-making.[73]

Transparency in data collection and management remains insufficient in the EU, leading to sanctions against non-compliant entities.[74] GDPR mandates clear and accessible information for data subjects. Indeed, the GDPR puts forward several principles including the transparency principle and accountability. Nevertheless, in cases like facial recognition, it's often hard to identify collectors or understand data use. Lack of transparency means individuals cannot control their data's processing and enforce their data subject rights.

Challenges to enforce GDPR with foreign companies

As reminded in the Clearview case, the GDPR territorial scope is quite broad and applies to "a foreign entity not established in the EU but still targeting individuals located in the EU to offer them a service and/or monitor their behavior to allow their identification"[75]. However, if the Regulation applies to such a company, enforcement decision from national DPAs is not as straightforward when the company has no physical presence in the EU.[76]

The wide accessibility of facial recognition

In 2005 already, A. Acquisti's experiments highlighted the future risks of facial recognition technology.[77] Research findings suggest a future where anyone with a smartphone could recognize and potentially stalk individuals. The rapid technological evolution suggests that street-level identification might become commonplace. For instance, in 2021, UK police developed a real-time facial recognition app. The app aims to identify vulnerable, missing and

[72] 'Facial Recognition: 20 Million Euros Penalty against CLEARVIEW AI' <https://www.cnil.fr/en/facial-recognition-20-million-euros-penalty-against-clearview-ai> accessed 21 September 2023.
[73] Wojciech Wiewiórowski (n 52).
[74] Liberties.EU (n 57).
[75] Jasserand (n 54).
[76] Jasserand (n 69).
[77] Ralph Gross and Alessandro Acquisti, 'Information Revelation and Privacy in Online Social Networks' <https://papers.ssrn.com/abstract=4253049> accessed 20 September 2023.

wanted individuals.[78] Videos published on YouTube even demonstrate how individuals can create their own facial recognition systems for public identification.[79]

4.4.1 Facial recognition and the AI Act

The EC proposal released in April 2021 restricts facial recognition in public places unless it is to fight "serious" crime, such as kidnappings and terrorism.[80] In September 2021, the European Parliament called for a total ban on facial recognition.[81]

In May 2023, the EDPB adopted its final version of the Guidelines 05/2022 on the use of facial recognition technology in the area of law enforcement.[82] The guidelines emphasise that facial recognition tools should only be used if necessary and proportionate, as laid down in the Charter of Fundamental Rights and repeats its call for a ban on facial recognition in certain cases such as in publicly accessible areas. However, these guidelines are specific to the law enforcement sector.

As already mentioned, in June 2023, the European Parliament after two years of discussion finally adopted its position on the proposal text of the AI Act. The MEPs adopted amendments against the use of facial recognition in public places. The MEPs removed the exceptions foreseen in the EC proposal text to avoid any risks of mass surveillance. MEPs don't seem ready to let go on this point during the negotiation according to Politico.[83] At the time of writing this deliverable, the AI Act proposal is being discussed and negotiated in trilogues.

Ethical considerations

The EDPS underlines the fundamental ethical question beyond privacy concerns. Indeed, facial recognition raises an important question for a democratic society: "turning the human face into another object for measurement and categorisation by automated processes controlled by

[78] South Wales Police, 'New Facial Recognition Mobile App to Identify Vulnerable, Missing and Wanted Individuals' (7 December 2021) <https://www.south-wales.police.uk/news/south-wales/news/2021/december/new-facial-recognition-app-to-to-identify-wanted-individuals/> accessed 20 September 2023.

[79] Liberties.EU (n 57).

[80] European Commission, Proposal fora 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. 2021 [COM(2021) 206 final].

[81] European Parliament, 'Report on Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters' (13 July 2021) <https://www.europarl.europa.eu/doceo/document/A-9-2021-0232_EN.html> accessed 20 September 2023.

[82] 'Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement | European Data Protection Board' <https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en> accessed 18 September 2023.

[83] Gian Volpicelli, 'Forget ChatGPT: Facial Recognition Emerges as AI Rulebook's Make-or-Break Issue' *POLITICO* (14 June 2023) <https://www.politico.eu/article/facial-recognition-artificial-intelligence-act-ai-issue-european-parliament/> accessed 21 September 2023.

powerful companies and governments touches the right to human dignity - even without the threat of it being used as a tool for oppression by an authoritarian state."[84]

In addition, facial recognition violates other fundamental rights at the core heart of democracy. It includes of course personal privacy and freedom, eroding the comfort of moving freely without constant surveillance.[85] Moreover, pre-emptive surveillance based on potential future wrongdoing infringes upon the presumption of innocence, leading to unjust treatment.[86]

Facial recognition technologies also raise concerns about infringement on freedom of speech and association and how they can be used for widespread biometric surveillance, particularly during public demonstrations, limiting freedom of expression and stifling political activism.[87]

The tension between the need for security and other fundamental rights needs to be clearly balanced out. For instance, in 2020, a French bill entitled "global security law" seemed to "give disproportionate priority to security over fundamental freedoms"[88]. The bill enabled police officers wearing bodycams to access the images and transmit them in real-time to the command post. The combination of these real-time body-cam images with the criminal record file could preventively identify individuals during protests, confiscate equipment, or place them in preventive custody.[89] In addition, the images captured by drones could be analyzed by real-time facial recognition, facilitating targeted police intervention against previously identified activists.

These elements are of great concern for the protection of the right to association, freedom of expression, freedom to have political opinions. This technology also undermines the right to anonymity, a basic expectation even in public spaces. The fear of constant surveillance hinders people's activities and self-expression.

In addition, because of the biases and related error results, minorities and vulnerable groups in society will be the ones that will be the most impacted by facial recognition. Liberties.EU warns about the potential danger of creating normalization around facial recognition.

While facial recognition offers innovative solutions for security, convenience, and efficiency across various sectors, it also raises important concerns that demand our utmost attention and caution. The continuous surveillance and tracking capabilities of this technology present a fundamental challenge to personal privacy, raising questions about the boundaries of surveillance in our increasingly digitized world.

Furthermore, facial recognition systems are not immune to biases and inaccuracies, which can lead to discrimination and wrongful identification, particularly against marginalized

[84] Wojciech Wiewiórowski (n 52).

[85] Liberties.EU (n 57).

[86] Liberties.EU (n 57).

[87] Liberties.EU (n 57).

[88] Noémie Krack, 'Loi de Sécurité Globale : Are Fundamental Rights and the Rule of Law Put in Danger by the French Bill ? (Part I)' (*CITIP blog*, 3 December 2020) <https://www.law.kuleuven.be/citip/blog/loi-de-securite-globale-are-fundamental-rights-and-the-rule-of-law-put-in-danger-by-the-french-bill-part-i/> accessed 21 September 2023.

[89] Krack (n 88).

communities. The potential for algorithmic biases to reinforce existing inequalities is a critical ethical concern.

As we explore these challenges, it is crucial to exercise restraint and establish robust regulations and guidelines for the responsible use of facial recognition technology. Striking a balance between its benefits and potential risks requires careful consideration and adherence to ethical and GDPR legal principles: ethics by design, security, fairness, oversight, accountability, and review.

5 Legal and ethical considerations of using generative AI models

5.1 Introduction

Generative AI refers to "deep-learning models that can generate high-quality text, images, and other content based on the data they were trained on."[90] It allows to create text, pictures, videos, from just a few text, image, sound prompts. Very recently we have seen an explosion in the launch and use of such tools – think of ChatGPT[91], Midjourney[92] or DALL-E[93]. These AI tools have the potential to transform many areas, from the medical sector, to creating new search engine architectures or personalised therapy bots.[94] The key characteristics of generative AI – their large size, opacity and potential to develop unexpected capabilities beyond those intended by their producers – raise many questions concerning perpetuating stereotypes and social biases, use toxic language, providing false or misleading information, and generate harmful and criminal content. [95]

Since the massive explosion in use and misuse of generative AI, some have called to "pause" the developments of generative AI. In particular, in an open letter signed by Elon Musk, Gary Marcus, Steve Wozniak, and over 1,800 signatories, including engineers from Amazon, DeepMind, Google, Meta, and Microsoft, the authors "call on all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4."[96] Others have criticized the idea, pointing out to the misrepresentation of research results and lack of clarity on the scope of the pause.[97] They argued that "instead of halting development, we should focus on fostering responsible and ethical AI research, addressing immediate concerns like biases, and promoting transparency and collaboration within the AI community."[98] In Belgium, a

[90] 'What Is Generative AI?' (*IBM Research Blog*, 9 February 2021) <https://research.ibm.com/blog/what-is-generative-AI> accessed 27 September 2023.
[91] See OpenAI, 'Introducing ChatGpt'<https://openai.com/blog/chatgpt> accessed 5 April 2023.
[92] 'Midjourney' <https://www.midjourney.com/home/?callbackUrl=%2Fapp%2F> accessed 5 April 2023.
[93] OpenAI, 'DALL-E2', <https://openai.com/product/dall-e-2> accessed 5 April 2023.
[94] EPRS, 'General-purpose artificial intelligence', available at: https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf
[95] Ibid.
[96] 'Pause Giant AI Experiments: An Open Letter' (*Future of Life Institute*) <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> accessed 27 September 2023.
[97] Koza Kurumlu, 'Why the 6-Month AI Pause Is a Bad Idea' (*Medium*, 8 April 2023) <https://medium.com/@koza.kurumlu/why-the-6-month-ai-pause-is-a-bad-idea-a6447123c346> accessed 27 September 2023.
[98] Kurumlu (n 97).

discussion around human-imitating AI was sparked by the chatbot-incited suicide.[99] In an "Open Letter: We are not ready for manipulative AI – urgent need for action" researchers have called to urgently set up awareness campaigns that better inform people of the risks associated with AI systems, more investment into research on AI's impact on fundamental rights, including the right to physical and moral integrity and regulation. [100]

In the media sector, generative AI can be used in a meaningful way to illustrate news stories, simplify text for different audiences, summarize documents or write potential headlines, or even to explore new angles or potential avenues for reporting.[101] Many newsrooms have already provided statements or guidelines[102] describing their approach to generative AI or even published articles written by ChatGPT.[103] But generative AI may also have much greater impact on the sector as the "roles and practices of creators and shifting the aesthetics of contemporary media".[104] With generative AI it is also becoming harder to distinguish between human and machine produced content.

Using generative AI means relying on training data made by a third party. This challenges the traditional concepts of the creative process and raises new ethical and legal challenges concerning accuracy, bias, privacy, data protection and intellectual property rights regarding generative AI systems. The below sub-sections will present a selection of identified challenges.

5.2 Data protection considerations

5.2.1 Types of data involved and the GDPR requirements

When considering data protection considerations, first off, one has to ask: what data is involved and does it involve the processing of personal data? Generally speaking, there are three broad categories of data involved in generative AI: (i) training data which is used by the developer to

[99] The Brussels Times, 'Belgian Man Dies by Suicide Following Exchanges with Chatbot' <https://www.brusselstimes.com/belgium/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt> accessed 27 September 2023.

[100] 'Open Letter: We Are Not Ready for Manipulative AI – Urgent Need for Action' <https://www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai> accessed 27 September 2023.

[101] See also: Generative AI In the Newsroom, <https://generative-ai-newsroom.com/> accessed 5 April 2023.

[102] Hannes Cools, 'Towards Guidelines for Guidelines on the Use of Generative AI in Newsrooms | by Hannes Cools | Generative AI in the Newsroom' <https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960> accessed 27 September 2023.

[103] Ryan Ermey, 'ChatGPT Wrote Part of This Article—It Didn't Go Great' <https://www.cnbc.com/2023/01/26/chatgpt-wrote-part-of-this-article-it-didnt-go-great.html> accessed 27 September 2023.

[104] Ziv Epstein, Aaron Hertzmann, and THE INVESTIGATORS OF HUMAN CREATIVITY, 'Art and the Science of Generative AI' (2023) 380 Science 1110.

train the model; (ii) input data provided by a user in a prompt; (iii) output data: a response given by a system (Figure 10).

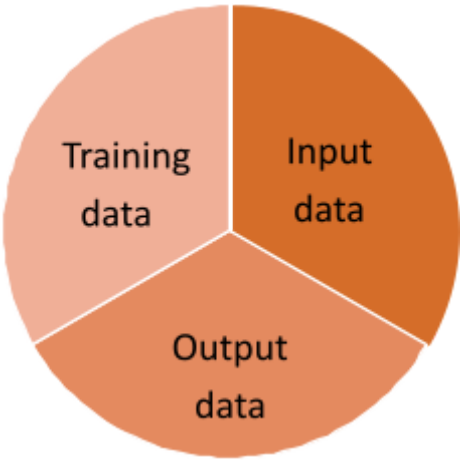


Figure 10 The overview of data involved in generative AI

Each of these categories may involve processing of personal data:

- training dataset may contain personal information;
- a user can provide personal data in a prompt; and
- an output data can also contain personal data, as also data derived/inferred from the input data can be personal data.

The data protection considerations about **training data** used for AI has already been subject to in-detail analysis in D4.3.

Personal data contained in a prompt adds additional complexity to the picture. It suffices to say that at the moment of writing, Open AI ChatGPT has around 100 million monthly users and Google Bard has around 30 million visits.[105] These millions of people use ChatGPT for various purposes but all do so by inserting prompts in the chat bar, prompts which may contain or reveal personal data (e.g. asking for details about herself).

As regards the **input** data, it is the responsibility of the user not to reveal personal data about other individuals. Output data is most likely to be the most problematic under the GDPR. It is worth reminding that Art. 5 of the GDPR contains a set of general principles, including accuracy and fairness. Controllers (AI developers, and providers), must ensure that personal data is accurate and avoid discrimination. It is well established that generative AI may provide

[105] Mark, 'Google Bard Statistics & Facts [July 2023]' (*MLYearning*, 18 July 2023) <https://www.mlyearning.org/google-bard-statistics-facts/> accessed 18 September 2023; Fabian Duarte, 'Number of ChatGPT Users (2023)' (*Exploding Topics*, 30 March 2023) <https://explodingtopics.com/blog/chatgpt-users> accessed 18 September 2023.

inaccurate or self-created responses and reflect bias contained within training data.[106] An interesting study by Luhang Sun et al. shows how DALL-E 2 underrepresents women in male-dominated fields and other representational and presentational biases in DALL-E 2.[107]

5.2.2 Data protection principles

While the generative AI technology is novel, the principles of data protection remain the same. In short, **organisations** that are developing or using generative AI should be able to answer the following questions:

- **What is a lawful basis for processing personal data?**

  Data collected to train the generative AI systems are scraped from the Internet. For example, Common Crawl data is 60% of the training data used in GPT-3; 50% of PaLM's training dataset is social media conversations; and OpenAI and Google have extensively used Reddit user posts in their large language models.[108] As a side note, as part of the new terms, Reddit now "reserves the right to charge fees for access and use of Reddit Services and Data, rates to be determined at Reddit's sole discretion."[109] The decision raises contractual law, consumer law and competition law (see Section 5.4.3) questions, as well as ethical concerns whether users' conversations should be monetized.

  Moreover, individuals are not aware that their data has been used to train these powerful AI systems. This raises difficulties as regards the lawful basis. Consent cannot be used, as it is impossible to ask millions of people for their agreement; legitimate interest can be used but a balancing exercise between the provider's (commercial) interest and data subject's right to privacy and data protection needs to be performed and documented. For the moment, it is extremely debatable whether the development of the AI systems or generative AI overrides the interest of the data subject's rights (see below an Italian DPA decision). The issue is even more problematic should a Large Language Model (LLM) be trained and retrained on special categories of personal data, subject to the strict rule of Article 9. Users of these systems may, through their prompts, reveal sensitive data (concerning them or third parties).[110] In principle, the legitimate interest of the controller cannot constitute a

[106] Luhang Sun and others, 'Smiling Women Pitching Down: Auditing Representational and Presentational Gender Biases in Image Generative AI' <http://arxiv.org/abs/2305.10566> accessed 27 September 2023; Doraid Dalalah and Osama MA Dalalah, 'The False Positives and False Negatives of Generative AI Detection Tools in Education and Academic Research: The Case of ChatGPT' (2023) 21 The International Journal of Management Education 100822; 'Generative AI Marks the Beginning of a New Era for Disinformation' (*EDMO*) <https://edmo.eu/2023/04/05/generative-ai-marks-the-beginning-of-a-new-era-for-disinformation/> accessed 27 September 2023.

[107] Sun and others (n 106).

[108] Dawen Zhang and others, 'Right to Be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' <http://arxiv.org/abs/2307.03941> accessed 21 September 2023.

[109] 'Data API Terms - Reddit' <https://www.redditinc.com/policies/data-api-terms> accessed 27 September 2023.

[110] Sophie Stalla-Bourdillon and Pablo Trigo Kramcsák, 'ChatGPT and Lawful Bases for Training AI: A Blended Approach? — The Digital Constitutionalist' (18 July 2023) <https://digi-con.org/chatgpt-and-lawful-bases-for-training-ai-a-blended-approach/> accessed 21 September 2023.

ground constituting an exception to the general prohibition of processing of sensitive personal data in art. 9(1) GDPR.

- **How do I ensure transparency?**
  There are transparency issues when it comes to the data collected to train the generative AI systems. As a general principle (see above Art. 12-14 GDPR) you must make information about the processing publicly accessible unless an exemption applies. If it does not take disproportionate effort, you must communicate this information directly to the individuals the data relates to.

- **How will I mitigate security risks?**
  You should apply security measures to eliminate and minimize the risks of data breaches and other interferences, such as data poisoning and other forms of adversarial attacks.

- **How do I keep my processing to minimum?**
  As a general principle, you must collect only the data that is adequate to fulfil your stated purpose. The data should be relevant and limited to what is necessary. In practice, however, with massive training data sets being used to train LLMs, this principle is difficult to achieve. To tackle the issue, experts have called for a significant time to be spent on assembling datasets suited for the tasks at hand rather than ingesting massive amounts of data from convenient or easily-scraped Internet sources.[111]

- **How will I comply with individual rights requests?**
  You must be able to respond to people's requests for access, rectification, erasure or other information rights. The overall problem of what happens should an individual request that their data is deleted from the training model, remains applicable to the generative AI. Researchers argue that because of the way LLMs store, and process information, they pose new challenges for compliance with data subject rights.[112] First, in LLMs, it is hard to know what personal data are used in training and how to attribute these data to particular individuals. Second, it is difficult to remove data from a trained model, as model weights are a complex integration of the whole collection of training data. Moreover, rectifying data from models is a difficult task.[113]

  It should be mentioned that ChatGPT for example, now introduces data controls. The data controls offer the user/data subject the ability to turn off chat history and opt out from the user's conversations to be used to train OpenAI models.[114] A separate issue is user's awareness of such a functionality, as it requires users to go into the settings.

[111] Bender and others (n 32).
[112] Zhang and others (n 108).
[113] Zhang and others (n 108).
[114] 'Data Controls FAQ | OpenAI Help Center' <https://help.openai.com/en/articles/7730893-data-controls-faq> accessed 18 September 2023.

- **Will you use generative AI to make solely automated decisions which have legal or similarly significant effects?**
  If so, you must comply with data subject's further rights under Article 22 GDPR (see D4.3 for further details).

- **How do I ensure model accuracy?**
  Accuracy of generative AI models is yet another problematic issue. Researchers from Stanford and UC Berkeley show that ChatGPT's behavior has noticeably deteriorated over time.[115] The study points out that substantial LLM drifts emerge on the simple mathematical task: "GPT-4's accuracy dropped from 84.0% in March to 51.1% in June". [116] On the other hand, there was a large improvement of GPT-3.5's accuracy, from 49.6% to 76.2%."[117] The varying levels of quality and accuracy, remain a problem in light of the consistent compliance with the GDPR accuracy principle.

5.2.3 Italian ChatGPT ban

To better illustrate the problem of compliance with the GDPR principles, it is worth turning the gaze to Italy. In March 2023, the Italian Data Protection Authority (DPA) ordered OpenAI to temporarily stop processing personal data of Italian data subjects, over data protection considerations.[118] The decision, which sparked a vivid discussion and controversies, neither permanently banned nor censored ChatGPT; nor did it express a negative opinion against the use of AI as such. The decision was a response to alleged violations of specific interests and rights granted by the GDPR. The four main alleged breaches of the GDPR were:

i. the lack of providing data subjects with information about their rights;
ii. the lack of a suitable lawful ground for personal data processing;
iii. a violation of the data accuracy principle; and
iv. a violation of the rules protecting minors.

Amongst these four alleged violations, only one was directly related to the AI element – i.e., the lack of suitable lawful ground for processing personal data to train the algorithm.[119] This is the issue already discussed in D4.3. First, it is worth reminding that the fact that personal data being publicly accessible does not grant the data controller a general authorization to re-use and

[115] Lingjiao Chen, Matei Zaharia and James Zou, 'How Is ChatGPT's Behavior Changing over Time?' <http://arxiv.org/abs/2307.09009> accessed 21 September 2023.
[116] Chen, Zaharia and Zou (n 115).
[117] Chen, Zaharia and Zou (n 115).
[118] Measure of March 30, 2023 [9870832], available at: <https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870832>
[119] Lorenzo Gugliotta, 'ChatGPT's Data Protection "Saga": Our Opportunity to Rediscover the Social Grounding of the Law' (*CITIP blog*, 19 May 2023) <https://www.law.kuleuven.be/citip/blog/chatgpts-data-protection-saga-our-opportunity-to-rediscover-the-social-grounding-of-the-law/> accessed 22 August 2023. Gugliotta.

further process. Second, the developers can rely on a "legitimate interest" ground as a legal basis for data processing. Legitimate interests may match the purpose of building AI training datasets better than individuals' consent, given the quality, quantity, and relevance demands of data curation for AI training.[120] However, to rely on this legal basis the data controller must pass a "legitimate interest" test, taking into account the expectations and fundamental rights and interests of the data subjects. However, as the Italian case shows, this balancing exercise is far from trivial. As summed up by Vale, developers of large generative AI models need further clarity about whether and to what extent they can rely on the legitimate interests, lawful ground or other alternatives to consent when training their models".[121]

The order, although unexpected for some, was only temporary. After OpenAI has implemented safeguards to bring the ChatGPT processing in line with the GDPR, the Italian DPA has withdrawn its initial ban.[122] In particular, Italian DPA welcomed the measures implemented by OpenAI, such as:

- drafting and publishing, on its website, an information notice addressed to users and non-users, describing which personal data are processed under which arrangements for training algorithms, and recalling that everyone has the right to opt-out from such processing;
- expanding its privacy policy for users and made it also accessible from the sign-up page prior to registration with the service;
- granting all individuals in Europe, including non-users, the right to opt-out from processing of their data for training of algorithms also by way of an online, easily accessible ad-hoc form;
- introducing a welcome back page containing links to the new privacy policy and the information notice on the processing of personal data for training algorithms;
- introducing mechanisms to enable data subjects to obtain erasure of information that is considered inaccurate, whilst stating that it is technically impossible, as of now, to rectify inaccuracies;
- clarifying in the information notice for users that it would keep on processing certain personal data to enable performance of its services on a contractual basis, however it would process users' personal data for training algorithms on the legal basis of its legitimate interest, without prejudice to users' right to opt-out from such processing;

[120] T. Kramcsak, 'Can Legitimate Interest Be an Appropriate Lawful Basis for Processing Artificial Intelligence Training Datasets? - ScienceDirect' <https://www-sciencedirect-com.kuleuven.e-bronnen.be/science/article/pii/S026736492200108X> accessed 22 August 2023.

[121] Sebastião Barros Vale, 'Training Large Generative AI Models Based on Publicly Available Personal Data: A GDPR Conundrum That the AI Act Could Solve — The Digital Constitutionalist' (14 April 2023) <https://digi-con.org/training-large-generative-ai-models-based-on-publicly-available-personal-data-a-gdpr-conundrum-that-the-ai-act-could-solve/> accessed 22 August 2023.

[122] ChatGPT: OpenAI reinstates service in Italy with enhanced transparency and rights for european users and non-users; available at: <https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9881490#english>

- implementing a form to enable all European users to opt-out from the processing of their personal data and thus to filter out their chats and chat history from the data used for training algorithms;
- adding a button to confirm that users are aged above 18 prior to gaining access to the service, or that they are aged above 13 and have obtained consent from their parents or guardians for that purpose;
- including the request to specify one's birthdate in the service sign-up page to block access by users aged below 13 and to request confirmation of the consent given by parents or guardians for users aged between 13 and 18.[123]

The Italian DPA will, however, continue to monitor the developments, in particular, whether OpenAI has introduced an age verification system and planning and conducted an information campaign to inform Italians of what happened as well as of their right to opt-out from the processing of their personal data for training algorithms. Moreover, other EU countries, such as Poland, have recently opened their own investigations.[124]

5.2.4 Chat GPT taskforce

Moreover, the EDPB announced in April 2023, the launch of a dedicated task force on ChatGPT.[125] This initiative follows the Italian DPA's decision to temporarily ban ChatGPT in Italy for infringement to data protection rules. Following this decision, other national DPA started investigations on privacy concerns of ChatGPT. The purpose of this task force is to align the DPA's approach to ChatGPT and ensure a consistent application of the rules, improve the cooperation and exchange of information on possible enforcement actions.

5.2.5 G7 data protection authorities' statement

It is also worth mentioning that the G7 data protection and privacy authorities issued a joint statement on generative AI. The statement states that drawing close communication between data protection and privacy authorities and companies at the design stage would mitigate or even eliminate some of the privacy risks. Generative AI providers to implement measures ensuring individuals can access, rectify and erase personal information.[126] Measures and tools to allow the exercise of privacy rights shall be provided to all data subjects.

---

[123] Ibid.

[124] ' Technologia musi być zgodna z RODO' <https://uodo.gov.pl/pl/138/2823> accessed 27 September 2023.

[125] Luca Bertuzzi, 'European Data Protection Authorities Launch Task Force on ChatGPT' *www.euractiv.com* (13 April 2023) <https://www.euractiv.com/section/artificial-intelligence/news/european-data-protection-authorities-launch-task-force-on-chatgpt/> accessed 15 September 2023.

[126] Office of the Privacy Commissioner of Canada, 'G7 Data Protection and Privacy Authorities Issue a Joint Statement on Generative AI Following Discussion on Emerging Technologies - Office of the Privacy

5.3 AI Act and foundation models

Definition of a foundation model

As already mentioned, the EP amendments to the AI Act propose to regulate "foundation models". As explained in Recital 60e, "*foundation models are a recent development, in which AI models are developed from algorithms designed to optimize for generality and versatility of output. Those models are often trained on a broad range of data sources and large amounts of data to accomplish a wide range of downstream tasks, including some for which they were not specifically developed and trained. The foundation model can be unimodal or multimodal, trained through various methods such as supervised learning or reinforced learning. AI systems with specific intended purpose or general purpose AI systems can be an implementation of a foundation model, which means that each foundation model can be reused in countless downstream AI or general purpose AI systems. These models hold growing importance to many downstream applications and systems.*"

To this end, new Article 3(1)(1c) defines **'foundation model'** as an AI system model that is trained on broad data at scale, is designed for generality of output, and can be adapted to a wide range of distinctive tasks. Pre-trained models developed for a narrower, less general, more limited set of applications that cannot be adapted for a wide range of tasks such as simple multi- purpose AI systems should not be considered foundation models (Recital 60g).

The definition and related obligations (see below) has been subject to criticism: "the current Article 28 b would symmetrically apply all obligations to every foundation model provider, both large and small."[127] It has been argued, that just like the DSA provides a specific set of rules for the very large online platforms, "the EU co-legislators could enact a similar approach to define Systemic Foundation Models in the AI Act".[128] This would help to narrow down the providers of foundation models to include only those who are systemically relevant. Others have argued that even small firms might produce foundation models which are then integrated into the applications and products which constitute high-risk uses of AI. For these reasons the principles of risk identification, testing and documentation should perhaps apply to all foundation models' providers, including non-systemic foundation models, but the granularity of testing and verification should be different.[129] "This kind of differentiation would allow for a more gradual and dynamic regulatory approach to foundation models".[130]

Commissioner of Canada' (22 June 2023) <https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230622_g7/> accessed 18 September 2023.

[127] 'A Law for Foundation Models: The EU AI Act Can Improve Regulation for Fairer Competition - OECD.AI' <https://oecd.ai/en/wonk/foundation-models-eu-ai-act-fairer-competition> accessed 21 August 2023. [128] 'A Law for Foundation Models: The EU AI Act Can Improve Regulation for Fairer Competition - OECD.AI' (n 127).

[129] 'Adapting the European Union AI Act to Deal with Generative Artificial Intelligence' (*Bruegel | The Brussels-based economic think tank*, 20 July 2023) <https://www.bruegel.org/analysis/adapting-european-union-ai-act-deal-generative-artificial-intelligence> accessed 21 August 2023.

[130] 'Adapting the European Union AI Act to Deal with Generative Artificial Intelligence' (n 129).

On the other hand, civil society points out that "defining foundation models as models "trained on broad data at scale" might focus too narrowly on data as one key variable and neglect the role of model size (i.e., the number of parameters in a model) and, as a function of these two, the compute used to train such a model."[131] It is in particular unclear what would qualify as "broad data at scale".

Obligations for the provider of a foundation model

Article 28b proposes to impose new obligations for the provider of a foundation model before making it available on the market or putting it into service, regardless of whether a foundation model is provided as a standalone model or embedded in an AI system or a product, or provided under free and open source licences, as a service, as well as other distribution channels. The provider of a foundation model shall:

1) demonstrate through appropriate design, testing and analysis the identification, the reduction and mitigation of reasonably foreseeable risks to health, safety, fundamental rights, the environment and democracy and the rule of law prior and throughout development with appropriate methods such as with the involvement of independent experts, as well as the documentation of remaining non-mitigable risks after development;
2) process and incorporate only datasets that are subject to appropriate data governance measures for foundation models, in particular measures to examine the suitability of the data sources and possible biases and appropriate mitigation;
3) design and develop the foundation model in order to achieve throughout its lifecycle appropriate levels of performance, predictability, interpretability, corrigibility, safety and cybersecurity assessed through appropriate methods such as model evaluation with the involvement of independent experts, documented analysis, and extensive testing during conceptualisation, design, and development;
4) design and develop the foundation model, making use of applicable standards to reduce energy use, resource use and waste, as well as to increase energy efficiency, and the overall efficiency of the system (Guidelines from the EC and standardization is expected on the practical implementation of this obligation);
5) design to enable the measurement and logging of the consumption of energy and resources, and, where technically feasible, other environmental impact the deployment and use of the systems may have over their entire lifecycle;
6) draw up extensive technical documentation and intelligible instructions for use, in order to enable the downstream providers to comply with their obligations;
7) establish a quality management system to ensure and document compliance;
8) register that foundation model in the EU database.

[131] 'The EU's AI Act and Foundation Models: Considerations for the Final Stretch of Negotiations' (*Mozilla Foundation*, 10 August 2023) <https://foundation.mozilla.org/en/blog/the-eus-ai-act-and-foundation-models-the-final-stretch/> accessed 21 August 2023.

Figure 11 presents the key requirements of Article 28b.

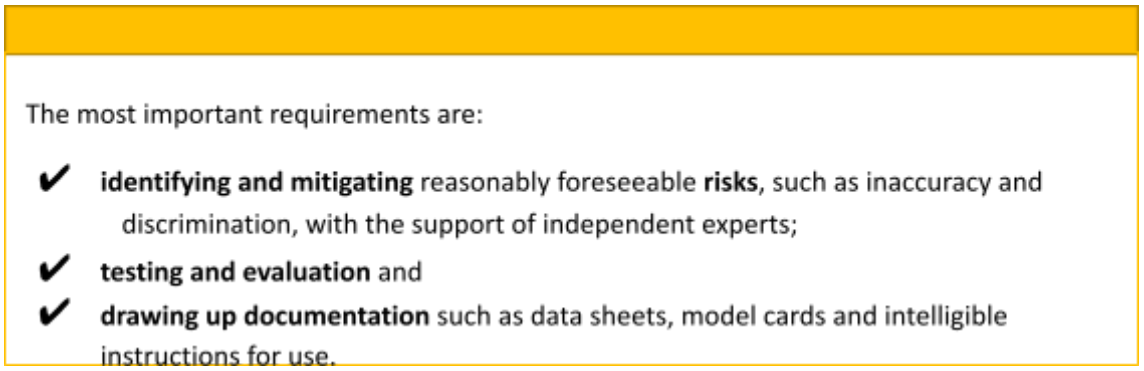


Figure 11 Key requirements of Article 28b

Annex VIII, Section C (new) provides a list of information which should be provided and kept up to date with regard to foundation models. The list includes contact details of the provider, trade name, a description of the data sources used in the development of the foundational model, the description of the capabilities and limitations of the foundation model, the description of the training resources used by the foundation model including computing power required, training time, description of the model's performance, description of the results of relevant internal and external testing and optimisation of the model.

The AI Office

Moreover, a new AI Office foreseen by the EP draft, could be equipped with the tasks of:

- providing monitoring of foundation models and organising a regular dialogue with the developers of foundation models with regard to their compliance as well as AI systems that make use of such AI models;
- providing interpretive guidance on how the AI Act applies to the ever evolving typology of AI value chains;
- providing particular oversight and monitoring and institutionalize regular dialogue with the providers of foundation models about the compliance of foundation models as well as AI systems that make use of such AI models with Article 28b of the Regulation, and about industry best practices for self-governance;
- issuing and periodically updating guidelines on the thresholds that qualify training a foundation model as a large training run, record and monitor known instances of large training runs, and issue an annual report on the state of play in the development, proliferation, and use of foundation models alongside policy options to address risks and opportunities specific to foundation models.[132]

It is also proposed that the European authorities on benchmarking and the AI Office, in close cooperation with international partners, jointly develop cost-effective guidance and capabilities

[132] Article 56b AI Act, EP draft.

to measure and benchmark aspects of AI systems and AI components, and in particular of foundation models.[133]

Generative AI

Moreover, providers of foundation models used in AI systems specifically intended to generate, with varying levels of autonomy, content such as complex text, images, audio, or video ("**generative AI**") and providers who specialise a foundation model into a generative AI system, shall in addition:

1) disclose that the content was generated and not provided by a human;
2) train, and where applicable, design and develop the foundation model in such a way as to ensure adequate safeguards against the generation of illegal content in line with the generally-acknowledged state of the art, and without prejudice to fundamental rights, including the freedom of expression;
3) document and make publicly available a sufficiently detailed summary of the use of training data protected under copyright law.

Feasibility of compliance

The analysis by Stanford University evaluated whether major foundation model providers currently comply with these AI Act requirements and found that they largely do not (Figure 12).[134] The analysis assessed 10 major foundation model providers (and their flagship models) for the 12 AI Act requirements. The best possible score is 48 as a result. The analysis shows that the "foundation model providers rarely disclose adequate information regarding the data, compute, and deployment of their models as well as the key characteristics of the models themselves. In particular, foundation model providers generally do not comply with draft

[133] Article 58a AI Act, EP draft.

[134] 'Stanford CRFM' <https://crfm.stanford.edu/2023/06/15/eu-ai-act.html> accessed 20 September 2023.

requirements to describe the use of copyrighted training data, the hardware used and emissions
produced in training, and how they evaluate and test models." [135]

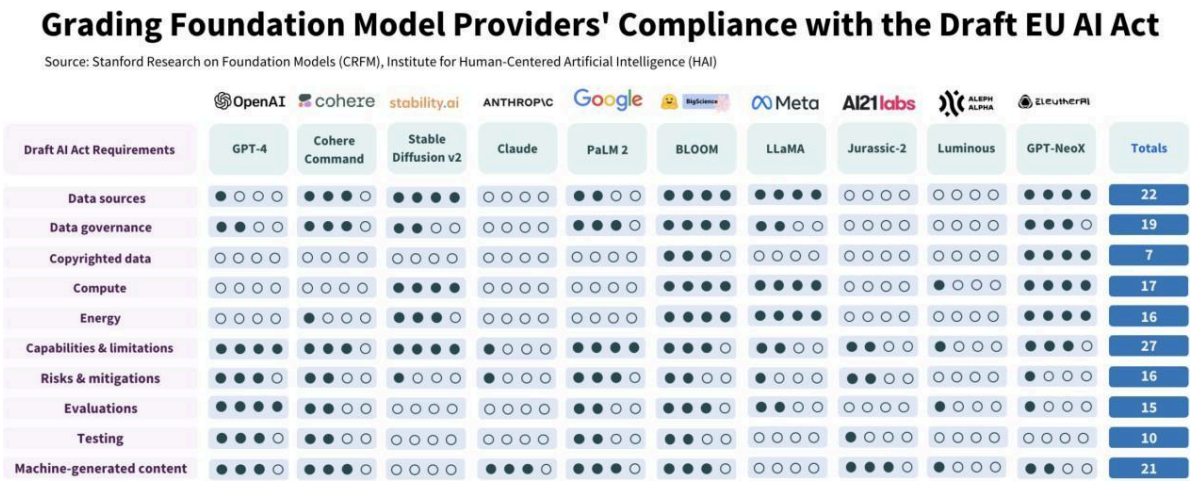


Figure 12 Foundation Model Providers' Compliance with the Draft EU AI Act

The report identifies the following persistent challenges:

- **Unclear liability due to copyright.** As few providers disclose any information about the copyright status of training data, the legality of using this data for training, compliance with specific licenses, and other intellectual property rights issues remains unclear (see also Section 5.4.2. on copyright)
- **Uneven reporting of energy use.** The report found that foundation model providers inconsistently report energy usage, emissions, and their mitigation measures to minimize emission (see also Section 5.5.6 on environmental effects).
- **Inadequate disclosure of risk mitigation/non-mitigation**. Few providers disclose the risks mitigations measures they implement and the efficacy of these mitigations. The Act also requires that providers describe "non-mitigated risks with an explanation on the reason why they cannot be mitigated", which none of the providers assessed do, the report concludes.
- **Absence of evaluation standards/auditing.** Foundation model providers rarely measure models' performance in terms of societal harms or factors such as robustness and calibration. Standards for foundation model evaluation still remain missing.

In conclusion, the study shows that there is a great room for improvement for most (if not all providers of the foundation models). The authors recommend that the "foundation model providers should work towards industry standards that will help the overall ecosystem become more transparent and accountable".[136] The Act (if enacted in the current form and enforced)

[135] 'Stanford CRFM' (n 134).

[136] 'Stanford CRFM' (n 134).

"would yield significant change to the ecosystem, making substantial progress towards more transparency and accountability."

5.4 Other legal considerations

5.4.1 DSA

As already mentioned in section 3.4, the DSA introduces new rules and obligations on the so-called intermediary services which cover Internet access providers, caching services, and "hosting" services, including online platforms and very large online platforms such as social media. Arguably, applications based on large generative AI models, such as ChatGPT do not qualify as either of these instances.[137] Chat GPT comes close to being consider a hosting service However, the definition of a hosting service requires the storage of information *provided by*, and at the request of, a user (Article 3(g)(iii) DSA). In the case of ChatGPT and alike, even though the input is prompted by the user, the output however, is not. This may suggest that there is a considerable risk that the DSA's new risk mitigation obligations and users' rights will not apply to generative AI models.

5.4.2 IP / Copyright

IP work needs to be protected from being duplicated and distributed freely with no compensation and control by those who invested the time, effort, and money to bring the work into existence. IP questions arise at the different stage of generative AI systems: training, input and output. There are numerous debates ongoing in the field of IP related to the lack of clarity in certain definitions and exceptions scope of the existing IP legal frameworks in force.

Training data

OpenAI stipulates that their "large language models are trained on a broad corpus of text that includes publicly available content, licensed content, and content generated by human reviewers."[138]. They also use the user's prompts and the data therein (unless they have opted out) to further train and improve their model.[139]

There are a lot of concerns and interrogations when it comes to the origin of data used to train these models. Is it part of the public domain or released under the most permissive licenses or data scrapped from the internet such as the LAION-5B dataset containing almost six billion tagged images compiled from scraping the web indiscriminately, and is known to include substantial number of copyrighted creations.

---

[137] Philipp Hacker, Andreas Engel and Theresa List, 'Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With input from ChatGPT' [2023] Verfassungsblog <https://verfassungsblog.de/chatgpt/> accessed 14 June 2023.

[138] 'Data Controls FAQ | OpenAI Help Center' (n 114).

[139] 'Data Controls FAQ | OpenAI Help Center' (n 114).

The IP legal framework is not absolute and can afford limitations. In the US there is a so-called fair use concept, whereas the EU offers a list of exceptions which can limit the protection granted. It includes the text and data mining (TDM) exception (Art. 3-4 EU Copyright Directive). It allows reproductions and extractions of lawfully accessible protected works for performing TDM, if it is made by research organizations and cultural heritage institutions. However, the exceptions can also be granted to other entities (including commercial entities such as OpenAI) provided that rightsholders have not expressly reserved this right. This requires an active supervision of IP holders of their IP catalogue and the need to expressly prohibit their IP work to fall under the text and data mining exception.

Even though TDM exception in principle permits the use of copyrighted material for training of foundation models, this provision does not appear in practice to have resolved the issue. Researchers have called to clarify the permitted uses of copyrighted content for training foundation models, and the conditions under which royalties must be paid.[140] Recent lawsuits about the use of copyrighted works for training Generative AI systems have underlined the need for legal certainty and clarity about the definition of the TDM exception.[141]

Some AI tools developed by Spawning helped artists opt-out over 1.4B images from public training datasets and creates an ai.txt file that selectively restricts or permits the use of artists' content for commercial AI training.[142] They can also identify and flag non-consenting data in datasets. However, this approach raises ethical concerns. It means that the responsibility is entirely on the IP rightsholders to investigate and waive their rights. They need to proactively look for their work in compiled or large-scale datasets. Some argue that it adds to the tension between market concentration in tech development and the asymmetry of power and information with copyright owners in light of the lack of transparency with large language models.[143] Others argue that it has the potential to increase the bargaining power of rights holders and lead to licensing deals with (and remuneration from) AI providers.[144]

[140] 'Adapting the European Union AI Act to Deal with Generative Artificial Intelligence' (n 129).

[141] Kyle Wiggers, 'The Current Legal Cases against Generative AI Are Just the Beginning' (*TechCrunch*, 27 January 2023) <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAE7FmNqEnZMTnfMqHtV8dq84sXKEadgyGvweOJcg6xcAjDCtjxkWmODAHljidNmTsC-GjYs8KyTTBO--rW2R_rAKHzvXVy3cKGJoIKWmGXQVJVs8GWPB-O_Y0cCkfVFzyNZM2BJEuW2uanUe7XbcAxlc1PHbPy35FpI59jBaXkZO&guccounter=2> accessed 18 September 2023.

[142] Spawning, 'AI Tools for Artists. Made by Artists.' <https://spawning.ai/> accessed 18 September 2023.

[143] João Pedro Quintais, 'A Primer and FAQ on Copyright Law and Generative AI for News Media' (*Medium*, 26 April 2023) <https://generative-ai-newsroom.com/a-primer-and-faq-on-copyright-law-and-generative-ai-for-news-media-f1349f514883> accessed 18 September 2023.

[144] João Pedro Quintais (n 143).

Recently, several media organisations[145] have decided to block OpenAI's GPTBot over their content.[146] GPTBot scrapes publicly accessible data online to feed into efforts to improve ChatGPT's accuracy which may include copyrighted material. OpenAI, mentioned that GPTBot will collect data from the entire internet except if the data are behind a paywall or from restricted sources.[147] The media organisation underlined that their investment in creative and trustworthy content need to be protected. They also mentioned that agreement with generative AI providers in the future are not excluded but that in the meantime their sites will block access to GPTbot. [148]

Input data

The terms of use of OpenAI state that users must comply with these Terms and all laws applicable. It adds that the user cannot use the service in a way that would harm the rights of third parties.[149] This means that "user input must not contain copyrighted data without the permission of the author or rights holder, or that the user must not use confidential data for no reason".[150] User input is not only used to generate output, but is also re-used to improve the service, and thus refine the AI model (unless the user has expressly opted out to this latest finality). The professional or business version of these generative AI systems and how they will handle the input data from users especially in a business context with confidential professional data needs to be assessed.

Output data

Generative AI models are creating a lot of discussions among IP scholars in the EU.[151] The qualification of their output raises questions. Is it an AI-assisted output or an AI-generated

[145] They include non-exclusively: The New York Times, CNN, Reuters, radio Frace, TF1, Les Echos, France24.

[146] Julia Tar, 'Several French Media Block OpenAI's GPTBot over Data Collection Concerns' (*www.euractiv.com*, 29 August 2023) <https://www.euractiv.com/section/artificial-intelligence/news/several-french-media-block-openais-gptbot-over-data-collection-concerns/> accessed 18 September 2023.

[147] 'GPTBot' (*OpenAI Platform*) <https://platform.openai.com> accessed 18 September 2023.

[148] Tar (n 146).

[149] 'Terms of Use' (March 2023) <https://openai.com/policies/terms-of-use> accessed 18 September 2023.

[150] Bernd Fiten and Edwin Jacobs, 'ChatGPT and Copyright: Many Questions Remain to Be Answered | Timelex' (*Timelex*, 23 March 2023) <https://www.timelex.eu/en/blog/chatgpt-and-copyright-many-questions-remain-be-answered> accessed 18 September 2023.

[151] A. Ramalho, "Will robots rule the artistic world? A proposed model for the legal status of creations by artificial intelligence systems", Journal of Internet Law 2017, https://ssrn.com/abstract=2987757 ; M. Senftleben, T. Margoni, D. Antal et al., "Ensuring the Visibility and Accessibility of European Creative Content on the World Market - The Need for Copyright Data Improvement in the Light of New Technologies and the Opportunity Arising from Article 17 of the CDSM Directive", Journal of Intellectual Property, Information Technology and E-Commerce Law, 12 February 2021, Vol. 13, No. 1, pp. 67-86, 2022, Available at SSRN: https://ssrn.com/abstract=3785272 or http://dx.doi.org/10.2139/ssrn.3785272 ; J. Vanherpe, "AI and IP : Great Expectations", Artificial Intelligence and the law, 2023, Intersentia, pp. 233-267. https://www.researchgate.net/publication/356811479_AI_and_IP_-_Great_Expectations

work?[152] Can an AI system be considered as an author under IP law? In the EU, the question of protection of AI output is not straightforward. The Court of Justice of the European Union confirmed that copyright protection requires originality including some form of human input to reflect the author's personality and the need for the work to express the author's own intellectual creation.[153] The whole copyright protection system is working around a human author. AI is not a legal entity, many sub-rights of copyright (economical or individual) would therefore not be attributable.

It remains to be seen whether introducing highly sophisticated and curated prompts in generative AI systems will be meeting the above mentioned requirements. Copyright protection can only be granted in case of a sufficient human intervention during the creative process.[154] The 'personal stamp must be discernible in the end result'.[155]

Conclusion on IP

The various litigations abroad against generative AI systems providers will surely provide more clarity on the topic. For the moment no litigation has been brought to the EU. João Pedro Quintais and Nick Diakopoulos, two scholars, wrote a FAQ on generative AI and copyright for

[152] European Commission, Directorate-General for Communications Networks, Content and Technology, Hartmann, C., Allan, J., Hugenholtz, P. et al., 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/10.2759/683128.

[153] CJEU 16 July 2009, n°C-5/08, Infopaq/Danske Dagblades Forening.

[154] Resolution 2019 – Study Question Copyright in artificially generated works, available at: "https://www.aippi.org/content/uploads/2022/11/Resolution_Copyright_in_artificially_generated_works_English.pdf"

[155] J. Vanherpe, 'AI and IP: Great Expectations', Artificial Intelligence and the law, 2023, Intersentia, pp. 261. https://www.researchgate.net/publication/356811479_AI_and_IP_-_Great_Expectations

news media.[156] The authors conclude that ethical and responsible use for media professionals including journalists is requested. They suggest the following a starting points (Figure 13):

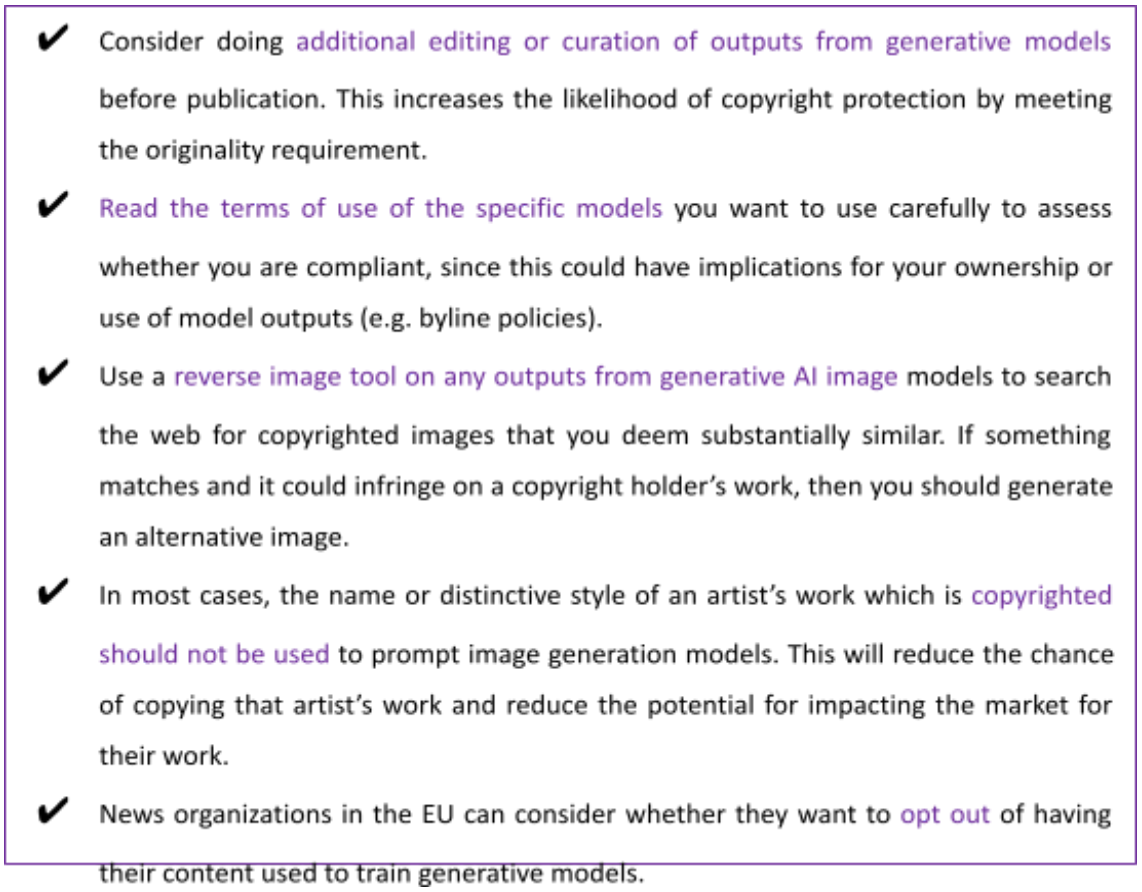


Figure 13 J. Quintais & N. Diakopoulos suggestions for media organisations to use responsibly generative AI systems.

5.4.3 Competition

The development of generative AI systems is often held by big tech companies: OpenAI, Microsoft, Google and Meta.[157] As a result, the ownership of data and models are often highly centralised, leading to market concentration and competition issues.[158] Few companies control the development of systems which will be used widely and globally. Indeed, "achieving state-of-the-art performance still requires a high budget, thus creating an entry barrier for potential

[156] J. Quintais & N. Diakopoulos, A Primer and FAQ on Copyright Law and Generative AI for News Media, (2023), https://infojustice.org/archives/45255

[157] Ayse Gizem Yasar and others, 'AI and the EU Digital Markets Act: Addressing the Risks of Bigness in Generative AI' <http://arxiv.org/abs/2308.02033> accessed 18 September 2023.

[158] Christophe Carugati, 'The Age of Competition in Generative Artificial Intelligence Has Begun' (*Bruegel*, 11 May 2023) <https://www.bruegel.org/first-glance/age-competition-generative-artificial-intelligence-has-begun> accessed 26 June 2023.

players and inhibiting diversity and market growth".[159] For instance, the GPT-4 Technical Report by OpenAI states that "given both the competitive landscape and the safety implications of large-scale models like GPT-4, this report contains no further details about the architecture (including model size), hardware, training compute, dataset construction, training method, or similar."[160]

EU competition rules have a broad scope; therefore, the generative AI systems can fall in the scope of the competition rules such as the prohibition of abuse of dominance (art. 102 TFEU) through the lack of transparency, (self)referencing techniques and excessive processing of users' data. The main characteristic of the EU competition rules is their *ex-post* approach, meaning that they can only play a role when a certain harmful behavior already took place. To solve this issue, the EU has adopted the Digital Market Act[161] also known as the DMA establishing *ex ante* obligation for big tech players.

The DMA establish new obligations for gatekeepers in order to solve the asymmetry of power and information between gatekeepers and other businesses or end-users. The gatekeepers are defined as tech companies that (a) act as gateways between consumers and businesses, and (b) provide "core platform services", namely services that have been included in a closed list in the DMA (e.g., app stores, video-sharing platforms). They must also enjoy a durable position in the market. On 6 September 2023, the EC unveiled the list designating the gatekeepers.[162] The gatekeepers designated are Alphabet (Google's parent company), Amazon, Apple, ByteDance (TikTok's parent company), Meta and Microsoft.

However, the DSA does not seem to be adapted to scope the generative AI systems providers. The list of core platform services does no match and does not explicitly mention generative AI systems services. The DMA would need to be revised for them to be covered.

5.5 Ethical considerations

Generative AI systems used in the media sector raise several concerns which have the potential to dramatically impact our society. Through our work on AI and media, we already explored and detailed horizontal AI issues in previous AI4Media deliverables: biases and discriminations, media independence, inequalities in access to AI, labour displacement, privacy, transparency, accountability and liability. We recommend reading D2.2 "*Initial White Paper on the social, economic, and political impact of media AI Technologies*" and D2.4 "*Pilot Policy*

[159] Yasar and others (n 157).

[160] OpenAI, 'GPT-4 Technical Report' <http://arxiv.org/abs/2303.08774> accessed 21 September 2023.

[161] Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance) 2022 (OJ L).

[162] Thierry Breton [@ThierryBreton], 'It's D-Day for #DMA! The Most Impactful Online Companies Will Now Have to Play by Our EU Rules. #Gatekeepers Are: Alphabet Amazon Apple ByteDance Meta Microsoft DMA Means More Choice for Consumers. Fewer Obstacles for Smaller Competitors. Opening the Gates to the Internet Https://T.Co/xaTluUfBax' <https://twitter.com/ThierryBreton/status/1699354391101260102> accessed 21 September 2023.

Recommendations for the use of AI in the Media Sector". These considerations also apply to generative AI systems. However, additional ones apply as well.

On September 13th 2023, AI4Media co-organised an online workshop with VISION, HumanE-AI-Net, ELISE, ELSA, euRobin and TAILOR. This second cross-cutting Theme Development Workshop was entitled "Trusted AI: The future of creating ethical and responsible AI systems". During the workshop, the participants had the opportunity to discuss with other experts the importance of ethical and responsible AI systems under the umbrella of Trusted AI. The report of this event will be made available soon.[163] A breakout session moderated by Noémie Krack from KU Leuven focused on "Ethical considerations and new challenges of Generative AI". The session included two expert talks. One from Dr Ana Maria Corrêa from KU Leuven on ethical and legal aspects of generative AI and one from Dr. Mark Schutera from ZF Group & KIT on the technical aspects. Certain aspects of the discussion are integrated below.

5.5.1 Manipulation and AI anthropomorphism

There is a risk with advanced large language models to deceive users. The phenomenon of AI anthropomorphism is a real issue which can lead to manipulation. AI anthropomorphism means that humans start to project human-like qualities or behaviour to non-human entities such as AI systems.[164] When AI systems start to sound like us, it can deceive people that AI systems could have ethical thinking or moral consideration or a responsibility or understanding of how they respond to the prompts. A dramatic example of this is when a man ended his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change.[165] The victim had started seeing the chatbot as a sentient being and the latter encouraged suicidal thoughts.

Generative AI's potential for emotional manipulation, such as using persuasive language or manipulating user emotions, raises high ethical concerns. However, while users might be deceived with high quality human like conversation, human oversight is considerably reduced with generative AI systems. The systems produce autonomously content without humans in the loop.[166]

5.5.2 Disinformation

For generative AI systems to function, they need a lot of information. Information which can be scrapped from the internet. However, harvesting data all over the internet includes mis- and

[163] The full report on the key findings from the workshop will be made available at https://www.vision4ai.eu/tdw-trusted-ai/.

[164] Chenhao Tan, 'On AI Anthropomorphism' (*Human-Centered AI*, 8 June 2023) <https://medium.com/human-centered-ai/on-ai-anthropomorphism-abff4cecc5ae> accessed 18 September 2023.

[165] Imane El Atillah, 'AI Chatbot Blamed for "encouraging" Young Father to Take His Own Life' (*euronews*, 31 March 2023) <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-> accessed 18 September 2023.

[166] David Sweenor, 'Generative AI Ethics' (*Medium*, 28 July 2023) <https://towardsdatascience.com/generative-ai-ethics-b2db92ecb909> accessed 21 September 2023.

disinformation content present online. In addition, generative AI systems enable the creation of increasingly complex deepfakes. Furthermore, they can also contribute to the spread of mis- and disinformation. When ChatGPT was asked to reply to the following simple question: "Could you please give me an example of a disinformation campaign in social media?", the system actually described in detail the various method steps to set up a disinformation campaign on social media. The generative system can serve both benign and malign requests. The consequence of using unreliable data is also illustrated by inaccurate responses for news queries from search engines using generative AI. Research into Bing's generative AI accuracy for news queries shows that there are detail errors, attribution errors, and the system also sometimes asserts the opposite of the truth.[167]

Moreover, a confident response by an AI system that does not seem to be justified by its training data is being referred to as AI hallucinations: "hallucination is the term employed for the phenomenon where AI algorithms and deep learning neural networks produce outputs that are not real, do not match any data the algorithm has been trained on, or any other identifiable pattern."[168]

Many users tend to forget that generative AI systems are not search engines. They are designated to generate ideas, concepts and narratives based on the data they were trained on and the user prompts. Generative AI systems, because they are capable of creating content have the potential to shape narratives, influence opinions, and even manipulate information. All of these elements risk eroding citizen's trust in information, public institutions, media and negatively impacting the right to receive information to have an informed public debate.

5.5.3 Creativity uniformisation

Generative AI systems have a dual role in the realm of creativity.[169] On the one hand, they have the potential to stimulate and enhance human creativity in several ways. AI-powered tools can assist human creativity by generating and suggesting creative content, and even helping individuals overcome creative blocks. They can analyse vast datasets to identify emerging trends, enabling content creators, artists, and writers to explore new and exciting topics. Additionally, AI can offer recommendations for content generation, helping creators fine-tune

[167] Nick Diakopoulos, 'Can We Trust Search Engines with Generative AI? A Closer Look at Bing's Accuracy for News Queries' *Medium* (17 February 2023) <https://medium.com/@ndiakopoulos/can-we-trust-search-engines-with-generative-ai-a-closer-look-at-bings-accuracy-for-news-queries-179467806bcc> accessed 5 April 2023.

[168] José Antonio Ribeiro Neto Zezinho, 'ChatGTP and the Generative AI Hallucinations' (*Medium*, 11 July 2023) <https://medium.com/chatgpt-learning/chatgtp-and-the-generative-ai-hallucinations-62feddc72369> accessed 21 September 2023.

[169] Steven Peters, 'Will AI Art Overtake Human Creativity, or Will It Inspire a New Era of Innovation?' (Medium, 27 June 2023) <https://medium.com/@spetersftl/will-the-sinister-potential-of-ai-art-overtake-human-creativity-or-will-it-inspire-a-new-era-of-90247d6c8e01>

their work and tailor it to specific audiences. Some speak about democratisation of creativity, as AI systems enable individuals with limited artistic skills to engage in the creative process.[170]

However, this collaboration between AI and human creativity also raises ethical questions. As AI-generated content becomes more prevalent, there is the risk that overreliance on AI systems for creative tasks may diminish the uniqueness and authenticity of human-generated content. In addition, there is a risk that creative works generated with heavy AI involvement may lack the depth, emotional resonance, and personal touch that are hallmarks of genuine human creativity.[171] This ethical concern also extends to questions of authorship, ownership, and the potential devaluation of human creative endeavours.[172]

In addition, the technology is able to mimic the voices, writing or creative styles of individuals, raising concerns about identity theft and impersonation, potentially damaging the rights of public figures, media personalities and artistic production.

Some authors nuance the impact that the use of AI can have on human creativity and art. They point that similar concerns emerged with the invention of the cameras.[173]

5.5.4 Absence of ethics by design

The importance of incorporating ethical considerations into the design phase of generative AI systems needs to be emphasized. Design choices can have significant ethical implications, and proactively integrating ethical principles into the design process is essential to mitigate potential issues. Ethics and responsibility lie on the human input through the whole life cycle of the AI systems design and responsibility needs to be clearly allocated.

The release of generative AI without due ethical processes and comprehensive assessment raises profound concerns. Such hasty deployment not only risks undermining the ethical principles that should guide technological advancement but also poses potential harm to society.

5.5.5 Accountability

Generative AI systems developers, providers, users, and end-users all play distinct roles, and delineating their responsibilities, particularly in the context of ethical use and potential consequences, is a critical task. Different liability questions arise, both regarding civil and criminal liability. For instance, when someone commits suicide following discussion and encouragement by a generative AI based chatbot, the question how to allocate the responsibilities would need to be further clarified.

---

[170] Peters (n 143).

[171] Peters (n 143).

[172] Jair Ribeiro, 'A Quick Reflection on Some Ethical Implications of Creative AI' (Medium, 21 January 2021) <https://towardsdatascience.com/a-quick-reflection-on-some-ethical-implications-of-creative-ai-adba63acdd47> accessed 21 September 2023.

[173] C Horton, Michael White and Sheena Iyengar, Will AI Art Devalue Human Creativity? (2023).

Some of the liability questions will be addressed in the EU forthcomings AI Act, AI Liability Directive and the revised product liability directive.[174] The proposal for an AI Liability Directive will apply to all categories of AI systems regardless of whether the relevant economic operator commercializing the products is established in the EU. These rules would apply to both AI system providers and users, offering favourable conditions for claimants, especially for High-Risk AI Systems.[175] They would allow for evidence disclosure by providers or users to support claims for damages. The burden of proof includes a presumption of causality in favour of claimants, with variations based on whether the AI system is High-Risk or not. The proposal to revise the product liability directive inserts new provisions addressing liability for products such as software (including artificial intelligence systems).[176] The purpose is to "ensure broader protection for damage caused by AI systems by alleviating the burden of proof in compensation claims pursued under national fault-based liability regimes".[177]

However, these initiatives are focusing on AI systems and not specifically on generative AI systems while these systems present unique challenges. It remains to be seen in the negotiation process, if some amendments would focus on generative AI.

5.5.6 Environmental impact of generative AI systems

At the Computer, Privacy and Data protection conference of 2023 (CPDP) attended by KUL, rare were the sessions focusing on the environmental impact of the development, deployment, application and end of life of technology. However, one panel organised by AlgorithmWatch's entitled "The global harms of powering artificial intelligence towards a sustainable future of data use and governance", focused on this aspect. It shed light on how little the environmental impact of technology is addressed by legislation and policy initiatives. Little research has yet been done due to a lack of data and access to information from manufacturers, developers and deployers. More research on this aspect and more transparency requirements are necessary.[178]

Researcher points that "While a single large AI model is not going to ruin the environment, if a thousand companies develop slightly different AI bots for different purposes, each used by

[174] European Commission Proposal fora 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. (n 98); European Commission, Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive) 2022 [COM(2022) 496 final] ; European Commission, Proposal for a directive of the European Parliament and of the Council on liability for defective products 2022 [COM(2022) 495].

[175] Julia Apostle, 'Modified Liability for AI: EU Review of AI Liability Rules' (25 May 2023) <https://www.orrick.com/en/Insights/2023/05/Modified-Liability-for-AI-EU-Review-of-AI-Liability-Rules> accessed 21 September 2023.

[176] Stefano De Luca and European Parliamentary Research Service, 'New Product Liability Directive' (May 2023) <https://www.europarl.europa.eu/thinktank/en/document/EPRS_BRI(2023)739341> accessed 21 September 2023.

[177] Stefano De Luca and European Parliamentary Research Service (n 150).

[178] Noémie Krack, 'Shall We Finally Talk about the Elephant in the Room? Zoom on the Unaddressed Environmental Impact of AI Systems.' (*MediaFutures*, 26 May 2023) <https://mediafutures.eu/shall-we-finally-talk-about-the-elephant-in-the-room-zoom-on-the-unaddressed-environmental-impact-of-ai-systems/> accessed 18 September 2023.

millions of customers, the energy use could become an issue."[179] This is a current concern with the various generative AI systems developments and competition between various generative AI services providers and developers.

[179] Kate Saenko, 'Is Generative AI Bad for the Environment? A Computer Scientist Explains the Carbon Footprint of ChatGPT and Its Cousins' (*The Conversation*, 23 May 2023) <http://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096> accessed 18 September 2023.

6 Conclusions

As this deliverable has shown, the GDPR has established a foundation for the legal framework which governs trustworthy AI in the EU. With the interpretations by the CJEU, the EDPB and EDPS guidelines and decisions, it establishes a rather clear set of requirements and obligations for data controllers, processors and data subjects. It is less clear how to square the terms like 'developers', providers of AI or third-party data providers for generative AI models within these categories. Moreover, the applicability of these principles to real-life scenarios, especially in the field of facial recognition, poses a considerable challenge as regards the compliance with data protection legislation. However, the GDPR is not the only applicable legal framework. The DGA, the DSA, DMA and the Data Act, although not explicitly regulating the AI are still relevant. The key piece of the puzzle, the AI Act is still being negotiated which makes it impossible to provide a final analysis of the future legal framework applicable to AI.

The GDPR and the AI Act proposal in its first version presented by the EC were drafted way before the massive boom of algorithms, apps and generative AI. As shown in section 5, there are numerous legal and ethical considerations of using generative AI models. These relate to a variety of personal data used (in a training dataset, in a prompt, in an output) and the difficulty with compliance with GDPR principles. What is a legal basis to train a large language model? How to comply with data subjects' request to delete her data from a training data set? How to ensure model accuracy? There are no easy answers to these questions, and as illustrated by the Italian Data Protection Authority decision ordering OpenAI to temporarily stop processing personal data, they may pose a significant challenge to the development and deployment of LLMs.

Moreover, it is likely that the foundation models will fall within the scope of the AI Act. As proposed by the MEPs, new obligations could apply to the providers of a foundation model before making it available on the market or putting it into service, regardless of whether a foundation model is provided as a standalone model or embedded in an AI system or a product, or provided under free and open source licences. These obligations could include, among others, identifying and mitigating reasonably foreseeable risks, such as inaccuracy and discrimination; testing and evaluation and drawing up documentation such as data sheets, model cards and intelligible instructions for use. It remains to be seen whether they make it to the final text. However, research shows that major foundation model providers currently do not comply with these AI Act requirements. A few providers disclose any information about the copyright status of training data, inconsistently report on energy usage and emissions, and few disclose the risks mitigations measures they implement.

Moreover, there are other legal considerations which need to be taken into account when developing LLMs. Those relate recent IP/copyright lawsuits against the use of copyrighted works for training generative AI systems, and competition law concerns regarding the asymmetry of power and information between the major providers and other businesses or end-users.

Generative AI systems are raising a lot of ethical questions regarding risks of manipulation and AI anthropomorphism, disinformation, creativity uniformization, lack of accountability and environmental impact of generative AI systems. Instead of imposing a pause or ban of their development, we opt for responsible regulation. As Bender et al. put it, LLMs developments should include "weighing the environmental and financial costs, investing resources into curating and carefully documenting datasets rather than ingesting everything on the web, carrying out pre-development exercises evaluating how the planned approach fits into research and development goals and supports stakeholder values, and encouraging research directions beyond ever larger language models."[180] Significant resource allocation should be put towards dataset curation and documentation practices as well as active inclusion of the communities underrepresented on the Internet. However, we agree that the responsibility for this lies not only with AI developers and providers, but also with national, European, and international governments and institutions who should adopt protective legal framework as soon as possible.[181]

Finally, the media sector plays a crucial role in safeguarding freedom of expression and the right to be informed. For that reason, it is of paramount importance that the use of generative systems is carefully considered. Their independence and accuracy will be key to ensure trustworthiness and trust in media services.

[180] Bender and others (n 32).

[181] 'Open Letter: We Are Not Ready for Manipulative AI – Urgent Need for Action' (n 100).

7 References

'A Law for Foundation Models: The EU AI Act Can Improve Regulation for Fairer Competition - OECD.AI' <https://oecd.ai/en/wonk/foundation-models-eu-ai-act-fairer-competition> accessed 21 August 2023

'About | European Data Protection Supervisor' (23 August 2023) <https://edps.europa.eu/about-edps_en> accessed 15 September 2023

'Adapting the European Union AI Act to Deal with Generative Artificial Intelligence' (*Bruegel | The Brussels-based economic think tank*, 20 July 2023) <https://www.bruegel.org/analysis/adapting-european-union-ai-act-deal-generative-artificial-intelligence> accessed 21 August 2023

'Aktualności - UODO' <https://uodo.gov.pl/pl/138/2823> accessed 27 September 2023

Bender EM and others, 'On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? ', *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (ACM 2021) <https://dl.acm.org/doi/10.1145/3442188.3445922> accessed 27 September 2023

Bernd Fiten and Edwin Jacobs, 'ChatGPT and Copyright: Many Questions Remain to Be Answered | Timelex' (*Timelex*, 23 March 2023) <https://www.timelex.eu/en/blog/chatgpt-and-copyright-many-questions-remain-be-answered> accessed 18 September 2023

Bertuzzi L, 'European Data Protection Authorities Launch Task Force on ChatGPT' *www.euractiv.com* (13 April 2023) <https://www.euractiv.com/section/artificial-intelligence/news/european-data-protection-authorities-launch-task-force-on-chatgpt/> accessed 15 September 2023

——, 'Data Act: EU Institutions Finalise Agreement on Industrial Data Law' (*www.euractiv.com*, 28 June 2023) <https://www.euractiv.com/section/data-privacy/news/data-act-eu-institutions-finalise-agreement-on-industrial-data-law/> accessed 15 September 2023

Brandom R, 'Huawei Worked on Facial Recognition System to Surveil Uighurs, New Report Claims' (*The Verge*, 8 December 2020) <https://www.theverge.com/2020/12/8/22163499/huawei-uighur-surveillance-facial-recognition-megvii-uyghur> accessed 20 September 2023

'Can Legitimate Interest Be an Appropriate Lawful Basis for Processing Artificial Intelligence Training Datasets? - ScienceDirect' <https://www-sciencedirect-com.kuleuven.e-bronnen.be/science/article/pii/S026736492200108X> accessed 22 August 2023

'ChatGPT Wrote Part of This Article—It Didn't Go Great' <https://www.cnbc.com/2023/01/26/chatgpt-wrote-part-of-this-article-it-didnt-go-great.html> accessed 27 September 2023

Chen L, Zaharia M and Zou J, 'How Is ChatGPT's Behavior Changing over Time?' <http://arxiv.org/abs/2307.09009> accessed 21 September 2023

Christophe Carugati, 'The Age of Competition in Generative Artificial Intelligence Has Begun' (*Bruegel*, 11 May 2023) <https://www.bruegel.org/first-glance/age-competition-generative-artificial-intelligence-has-begun> accessed 26 June 2023

Clifford D and Ausloos J, 'Data Protection and the Role of Fairness' (2018) 37 Yearbook of European Law 130

Dalalah D and Dalalah OMA, 'The False Positives and False Negatives of Generative AI Detection Tools in Education and Academic Research: The Case of ChatGPT' (2023) 21 The International Journal of Management Education 100822

'Data API Terms - Reddit' <https://www.redditinc.com/policies/data-api-terms> accessed 27 September 2023

'Data Controls FAQ | OpenAI Help Center' <https://help.openai.com/en/articles/7730893-data-controls-faq> accessed 18 September 2023

Ducuing C and Margoni T, 'Data Act Blog Post Series: Introduction' (*CITIP blog*, 21 April 2022) <https://www.law.kuleuven.be/citip/blog/data-act-blog-post-series-introduction/> accessed 15 September 2023

EDPB, 'EDPB Letter to the European Commission on Adapting Liability Rules to the Digital Age and Artificial Intelligence (AI)' (22 February 2022) <https://edpb.europa.eu/our-work-tools/our-documents/letters/edpb-letter-european-commission-adapting-liability-rules_en>

'EDPB | European Data Protection Board' <https://edpb.europa.eu/edpb_en> accessed 15 September 2023

Epstein Z, Hertzmann A, and THE INVESTIGATORS OF HUMAN CREATIVITY, 'Art and the Science of Generative AI' (2023) 380 Science 1110

European Commission, 'Data Act – Questions and Answers*' (*Europa*) <https://ec.europa.eu/commission/presscorner/detail/en/qanda_22_1114> accessed 15 September 2023

European Data Protection Board, 'Guidelines 8/2020 on the Targeting of Social Media Users | European Data Protection Board' (13 April 2021) <https://edpb.europa.eu/our-work-tools/our-documents/guidelines/guidelines-82020-targeting-social-media-users_en> accessed 21 September 2023

European Data Protection Supervisor, 'Reshaping the EDPS to Tackle Data Protection Challenges' (23 August 2023) <https://edps.europa.eu/press-publications/press-news/press-releases/2023/reshaping-edps-tackle-data-protection-challenges> accessed 15 September 2023

European Parliament, ‘Report on Artificial Intelligence in Criminal Law and Its Use by the Police and Judicial Authorities in Criminal Matters’ (13 July 2021) <https://www.europarl.europa.eu/doceo/document/A-9-2021-0232_EN.html> accessed 20 September 2023

‘Explainable Artificial Intelligence Needs Human Intelligence | European Data Protection Supervisor’ <https://edps.europa.eu/press-publications/press-news/blog/explainable-artificial-intelligence-needs-human-intelligence_en> accessed 21 September 2023

Fabian Duarte, ‘Number of ChatGPT Users (2023)’ (*Exploding Topics*, 30 March 2023) <https://explodingtopics.com/blog/chatgpt-users> accessed 18 September 2023

‘Facial Recognition: 20 Million Euros Penalty against CLEARVIEW AI’ <https://www.cnil.fr/en/facial-recognition-20-million-euros-penalty-against-clearview-ai> accessed 21 September 2023

Friedman A and others, ‘Privacy Aspects of Recommender Systems’ in Francesco Ricci, Lior Rokach and Bracha Shapira (eds), *Recommender Systems Handbook* (Springer US 2015) <https://link.springer.com/10.1007/978-1-4899-7637-6_19> accessed 18 August 2023

Gabriel I, ‘Toward a Theory of Justice for Artificial Intelligence’ (2022) 151 Daedalus 218

‘Generative AI Marks the Beginning of a New Era for Disinformation’ (*EDMO*) <https://edmo.eu/2023/04/05/generative-ai-marks-the-beginning-of-a-new-era-for-disinformation/> accessed 27 September 2023
‘GPTBot’ (*OpenAI Platform*) <https://platform.openai.com> accessed 18 September 2023
Gross R and Acquisti A, ‘Information Revelation and Privacy in Online Social Networks’ <https://papers.ssrn.com/abstract=4253049> accessed 20 September 2023

Gugliotta L, ‘ChatGPT’s Data Protection “Saga”: Our Opportunity to Rediscover the Social Grounding of the Law’ (*CITIP blog*, 19 May 2023) <https://www.law.kuleuven.be/citip/blog/chatgpts-data-protection-saga-our-opportunity-to-rediscover-the-social-grounding-of-the-law/> accessed 22 August 2023

‘Guidelines 05/2022 on the Use of Facial Recognition Technology in the Area of Law Enforcement | European Data Protection Board’ <https://edpb.europa.eu/our-work-tools/documents/public-consultations/2022/guidelines-052022-use-facial-recognition_en> accessed 18 September 2023

Hacker P, Engel A and List T, ‘Understanding and Regulating ChatGPT, and Other Large Generative AI Models: With input from ChatGPT’ [2023] Verfassungsblog <https://verfassungsblog.de/chatgpt/> accessed 14 June 2023

Hildebrandt M, ‘The Issue of Proxies and Choice Architectures. Why EU Law Matters for Recommender Systems’ (2022) 5 Frontiers in Artificial Intelligence <https://www.frontiersin.org/articles/10.3389/frai.2022.789076> accessed 18 August 2023

Hill K, 'The Secretive Company That Might End Privacy as We Know It' *The New York Times* (18 January 2020) <https://www.nytimes.com/2020/01/18/technology/clearview-privacy-facial-recognition.html> accessed 21 September 2023

Imane El Atillah, 'AI Chatbot Blamed for "encouraging" Young Father to Take His Own Life' (*euronews*, 31 March 2023) <https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-> accessed 18 September 2023

ipvideomarket, 'Huawei / Megvii Uyghur Alarms' (*IPVM*, 8 December 2020) <https://ipvm.com/reports/huawei-megvii-uygur> accessed 20 September 2023

Jasserand C, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part I)' (*CITIP blog*, 28 April 2022) <https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-i/> accessed 21 September 2023

——, 'Clearview AI: Illegally Collecting and Selling Our Faces in Total Impunity? (Part II)' (*CITIP blog*, 5 May 2022) <https://www.law.kuleuven.be/citip/blog/clearview-ai-illegally-collecting-and-selling-our-faces-in-total-impunity-part-ii/> accessed 21 September 2023

João Pedro Quintais, 'A Primer and FAQ on Copyright Law and Generative AI for News Media' (*Medium*, 26 April 2023) <https://generative-ai-newsroom.com/a-primer-and-faq-on-copyright-law-and-generative-ai-for-news-media-f1349f514883> accessed 18 September 2023

Krack N, 'Loi de Sécurité Globale : Are Fundamental Rights and the Rule of Law Put in Danger by the French Bill ? (Part I)' (*CITIP blog*, 3 December 2020) <https://www.law.kuleuven.be/citip/blog/loi-de-securite-globale-are-fundamental-rights-and-the-rule-of-law-put-in-danger-by-the-french-bill-part-i/> accessed 21 September 2023

Kurumlu K, 'Why the 6-Month AI Pause Is a Bad Idea' (*Medium*, 8 April 2023) <https://medium.com/@koza.kurumlu/why-the-6-month-ai-pause-is-a-bad-idea-a6447123c346> accessed 27 September 2023

Kyle Wiggers, 'The Current Legal Cases against Generative AI Are Just the Beginning' (*TechCrunch*, 27 January 2023) <https://techcrunch.com/2023/01/27/the-current-legal-cases-against-generative-ai-are-just-the-beginning/?guce_referrer=aHR0cHM6Ly93d3cuZ29vZ2xlLmNvbS8&guce_referrer_sig=AQAAAE7FmNqEnZMTnfMqHtV8dq84sXKEadgyGvweOJcg6xcAjDCtjxkWmODAHljidNmTsC-GjYs8KyTTBO--rW2R_rAKHzvXVy3cKGJoIKWmGXQVJVs8GWPB-O_Y0cCkfVFzyNZM2BJEuW2uanUe7XbcAxlc1PHbPy35FpI59jBaXkZO&guccounter=2> accessed 18 September 2023

Liberties.EU, '7 Biggest Privacy Concerns Around Facial Recognition Technology| LibertiesEU' (*Liberties.eu*, 25 October 2022) <https://www.liberties.eu/en/stories/facial-recognition-privacy-concerns/44518> accessed 20 September 2023

Mark, 'Google Bard Statistics & Facts [July 2023]' (*MLYearning*, 18 July 2023) <https://www.mlyearning.org/google-bard-statistics-facts/> accessed 18 September 2023

Milano S, Taddeo M and Floridi L, 'Recommender Systems and Their Ethical Challenges' (2020) 35 AI & SOCIETY 957

'Model Cards - OECD.AI' <https://oecd.ai/fr/catalogue/tools/model-cards> accessed 21 August 2023

Munn L, 'The Uselessness of AI Ethics' (2023) 3 AI and Ethics 869

Noémie Krack, 'Shall We Finally Talk about the Elephant in the Room? Zoom on the Unaddressed Environmental Impact of AI Systems.' (*MediaFutures*, 26 May 2023) <https://mediafutures.eu/shall-we-finally-talk-about-the-elephant-in-the-room-zoom-on-the-unaddressed-environmental-impact-of-ai-systems/> accessed 18 September 2023

Office of the Privacy Commissioner of Canada, 'G7 Data Protection and Privacy Authorities Issue a Joint Statement on Generative AI Following Discussion on Emerging Technologies - Office of the Privacy Commissioner of Canada' (22 June 2023) <https://www.priv.gc.ca/en/opc-news/news-and-announcements/2023/an_230622_g7/> accessed 18 September 2023

'Open Letter: We Are Not Ready for Manipulative AI – Urgent Need for Action' <https://www.law.kuleuven.be/ai-summer-school/open-brief/open-letter-manipulative-ai> accessed 27 September 2023

OpenAI, 'GPT-4 Technical Report' <http://arxiv.org/abs/2303.08774> accessed 21 September 2023

'Pause Giant AI Experiments: An Open Letter' (*Future of Life Institute*) <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> accessed 27 September 2023

Raposo VL, '(Do Not) Remember My Face: Uses of Facial Recognition Technology in Light of the General Data Protection Regulation' (2023) 32 Information & Communications Technology Law 45

Rocher L, Hendrickx JM and De Montjoye Y-A, 'Estimating the Success of Re-Identifications in Incomplete Datasets Using Generative Models' (2019) 10 Nature Communications 3069

Saenko K, 'Is Generative AI Bad for the Environment? A Computer Scientist Explains the Carbon Footprint of ChatGPT and Its Cousins' (*The Conversation*, 23 May 2023) <http://theconversation.com/is-generative-ai-bad-for-the-environment-a-computer-scientist-explains-the-carbon-footprint-of-chatgpt-and-its-cousins-204096> accessed 18 September 2023

Sartor G and others, *The Impact of the General Data Protection Regulation (GDPR) on Artificial Intelligence: Study* (2020) <http://www.europarl.europa.eu/RegData/etudes/STUD/2020/641530/EPRS_STU(2020)641530_EN.pdf> accessed 15 June 2021

Smits M and others, 'Values That Matter: Mediation Theory and Design for Values', *Research Perspectives in the era of transformations: Conference proceedings* (Academy for Design Innovation Management 2019) <https://research.utwente.nl/en/publications/values-that-matter-mediation-theory-and-design-for-values> accessed 21 September 2023

South Wales Police, 'New Facial Recognition Mobile App to Identify Vulnerable, Missing and Wanted Individuals' (7 December 2021) <https://www.south-wales.police.uk/news/south-wales/news/2021/december/new-facial-recognition-app-to-to-identify-wanted-individuals/> accessed 20 September 2023

Spawning, 'AI Tools for Artists. Made by Artists.' <https://spawning.ai/> accessed 18 September 2023

Stalla-Bourdillon S and Kramcsák PT, 'ChatGPT and Lawful Bases for Training AI: A Blended Approach? — The Digital Constitutionalist' (18 July 2023) <https://digi-con.org/chatgpt-and-lawful-bases-for-training-ai-a-blended-approach/> accessed 21 September 2023

'Stanford CRFM' <https://crfm.stanford.edu/2023/06/15/eu-ai-act.html> accessed 20 September 2023

Sun L and others, 'Smiling Women Pitching Down: Auditing Representational and Presentational Gender Biases in Image Generative AI' <http://arxiv.org/abs/2305.10566> accessed 27 September 2023

Sweenor D, 'Generative AI Ethics' (*Medium*, 28 July 2023) <https://towardsdatascience.com/generative-ai-ethics-b2db92ecb909> accessed 21 September 2023

Tan C, 'On AI Anthropomorphism' (*Human-Centered AI*, 8 June 2023) <https://medium.com/human-centered-ai/on-ai-anthropomorphism-abff4cecc5ae> accessed 18 September 2023

Tar J, 'Several French Media Block OpenAI's GPTBot over Data Collection Concerns' (*www.euractiv.com*, 29 August 2023) <https://www.euractiv.com/section/artificial-intelligence/news/several-french-media-block-openais-gptbot-over-data-collection-concerns/> accessed 18 September 2023

Teodora Lalova-Spinks and Daniela Spajić, 'The Broadening of the Right to Data Portability for Internet-of-Things Products in the Data Act: Who Does the Act Actually Empower? (Part I)' (*CITIP blog*, 16 June 2022) <https://www.law.kuleuven.be/citip/blog/the-broadening-of-the-right-to-data-portability-for-internet-of-things-products-in-the-data-act-part-i/> accessed 15 September 2023

'Terms of Use' (March 2023) <https://openai.com/policies/terms-of-use> accessed 18 September 2023

'The EU's AI Act and Foundation Models: Considerations for the Final Stretch of Negotiations' (*Mozilla Foundation*, 10 August 2023) <https://foundation.mozilla.org/en/blog/the-eus-ai-act-and-foundation-models-the-final-stretch/> accessed 21 August 2023

'The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research by Erin Kenneally, David Dittrich :: SSRN' <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2445102> accessed 21 August 2023

Thierry Breton [@ThierryBreton], 'It's D-Day for #DMA! The Most Impactful Online Companies Will Now Have to Play by Our EU Rules. #Gatekeepers Are: Alphabet Amazon Apple ByteDance Meta Microsoft DMA Means More Choice for Consumers. Fewer Obstacles for Smaller Competitors. Opening the Gates to the Internet Https://T.Co/xaTluUfBax' <https://twitter.com/ThierryBreton/status/1699354391101260102> accessed 21 September 2023

Times TB, 'Belgian Man Dies by Suicide Following Exchanges with Chatbot' <https://www.brusselstimes.com/belgium/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt> accessed 27 September 2023

'Towards Guidelines for Guidelines on the Use of Generative AI in Newsrooms | by Hannes Cools | Generative AI in the Newsroom' <https://generative-ai-newsroom.com/towards-guidelines-for-guidelines-on-the-use-of-generative-ai-in-newsrooms-55b0c2c1d960> accessed 27 September 2023

Vale SB, 'Training Large Generative AI Models Based on Publicly Available Personal Data: A GDPR Conundrum That the AI Act Could Solve — The Digital Constitutionalist' (14 April 2023) <https://digi-con.org/training-large-generative-ai-models-based-on-publicly-available-personal-data-a-gdpr-conundrum-that-the-ai-act-could-solve/> accessed 22 August 2023

Vardanyan L and Kocharyan H, 'The GDPR and the DGA Proposal: Are They in Controversial Relationship?' (2022) 9 European Studies 91

Volpicelli G, 'Forget ChatGPT: Facial Recognition Emerges as AI Rulebook's Make-or-Break Issue' *POLITICO* (14 June 2023) <https://www.politico.eu/article/facial-recognition-artificial-intelligence-act-ai-issue-european-parliament/> accessed 21 September 2023

Wachter S, Mittelstadt B and Russell C, 'Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR' <http://arxiv.org/abs/1711.00399> accessed 21 August 2023

'What Is Generative AI?' (*IBM Research Blog*, 9 February 2021) <https://research.ibm.com/blog/what-is-generative-AI> accessed 27 September 2023

Williams ML, Burnap P and Sloan L, 'Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation' (2017) 51 Sociology 1149

Wojciech Wiewiórowski, 'Facial Recognition: A Solution in Search of a Problem?' (*European Data Protection Supervisor*, 28 October 2019) <https://edps.europa.eu/press-publications/press-news/blog/facial-recognition-solution-search-problem> accessed 21 September 2023

Yasar AG and others, 'AI and the EU Digital Markets Act: Addressing the Risks of Bigness in Generative AI' <http://arxiv.org/abs/2308.02033> accessed 18 September 2023

Zezinho JARN, 'ChatGTP and the Generative AI Hallucinations' (*Medium*, 11 July 2023) <https://medium.com/chatgpt-learning/chatgtp-and-the-generative-ai-hallucinations-62feddc72369> accessed 21 September 2023

Zhang D and others, 'Right to Be Forgotten in the Era of Large Language Models: Implications, Challenges, and Solutions' <http://arxiv.org/abs/2307.03941> accessed 21 September 2023

Zuiderveen Borgesius FJ and others, 'Should We Worry about Filter Bubbles?' (2016) 5 Internet Policy Review <https://policyreview.info/node/401> accessed 21 September 2021

European Commission, Proposal fora 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. 2021 [COM(2021) 206 final]

European Data Protection Board, Guidelines 3/2019 on processing of personal data through video devices 2020

Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act) (Text with EEA relevance) 2022 (OJ L)

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