The Role of Artificial Intelligence in Processing and Generating New Data

An exploration of legal and policy challenges in open data ecosystems

data.europa.eu
The official portal for European data

European Commission
Directorate-General for Communications Networks, Content and Technology
Unit G.1 Data Policy and Innovation
Email: CNECT-G1@ec.europa.eu

data.europa.eu
Email: info@data.europa.eu

Author:
Hans Graux
Pieter Gryffroy
Magdalena Gad-Nowak
Liesa Boghaert

Last update: July 2024
https://data.europa.eu/

DISCLAIMER
The information and views set out in this publication are those of the author(s) and do not necessarily reflect the official opinion of the Commission. The Commission does not guarantee the accuracy of the data included in this study. Neither the Commission nor any person acting on the Commission's behalf may be held responsible for the use that may be made of the information contained herein.

Luxembourg: Publications Office of the European Union, 2024

ISBN: 978-92-78-44246-0 doi: 10.2830/412108 Catalogue number: OA-02-24-797-EN-N

Table of Contents

An introduction to the potential impact of artificial intelligence systems on open data ecosystems

1. Artificial intelligence and open data ecosystems

The general impact of artificial intelligence (AI) systems on businesses, governments and the global economy is currently a hot topic. This isn't surprising, considering that AI is believed to have the potential to bring about radical, unprecedented changes in the way people live and work.

The transformative potential of AI originates to a large extent from its ability to analyse data at scale, and to notice and internalise patterns and correlations in that data that humans (or fully deterministic algorithms) would struggle to identify. In simpler terms: modern AIs flourish especially if they can be trained on large volumes of data, and when they are used in relation to large volumes of data.

A highly visible example of this process is the current popularity of 'generative' AI systems (AISs), which are capable of generating seemingly new texts, images, videos or other data at the user's request. They do so by analysing patterns in large volumes of input data (pre-existing texts, images and videos), from which they then deduce common patterns. Thereafter, based on prompts from the users, they can generate new outputs that reproduce the characteristics of the input data. Generative AI chat systems have been broadly taken up by the market and allow fast text responses to be generated that can easily be mistaken for qualified human answers. Comparable systems exist for image and video outputs.

Because of these characteristics, there is an inherent close connection between AI and open data. Compared to other computing techniques, AIs have a remarkable ability to extract insights from large datasets and to produce useful new outputs; but to make them work effectively, substantial sets of accessible data, to be used as training material, are essential. The accessibility and free use of large volumes of data are two of the main characteristics of open data. In other words: open data ecosystems can become – and may already be – the source material that high performance AIs need.

For AI systems (AISs) to function properly, the following three critical factors, known as the three Vs, are necessary.

- Data volume – AI requires significant amounts of data to be trained on.
- Data variety – diverse data sources enhance AI capabilities and reduces the risk of biases.
- Data veracity – bad training data will result in bad performance, so data truthfulness is crucial. Reliable sources play a role in determining data quality.

Open data can help to satisfy these preconditions. While none of the three Vs are inherently present in every single open dataset, the breadth of data will help to satisfy the volume and variety requirements. Moreover, in the European open data community, the reliability of data sources will

help to satisfy the veracity requirement. In summary, open data ecosystems have the potential to help construct reliable AIs by providing a repository of usable training data; and inversely, the open data community can benefit from AIs by using them as a tool to trawl through large datasets and obtain insights that would otherwise not be readily apparent. In this way, the combination of AI and open data has the potential to revolutionise data ecosystems, enabling innovation and facilitating informed decision-making.

2. Problem statement and structure of this research paper

Despite these clear potential benefits, AIs can also be a source of new challenges from a legal and policy perspective. Problems can present themselves on both the **input** side (how AIs are created and trained) and the **output** side (how they are brought to market and how their impacts can be managed and controlled).

On the input side, there are many legal concerns in relation to how AIs obtain access to training materials and whether their use of that training material is lawful. When the training materials consist of human-made creative works, they are likely to be subject to **intellectual property rights**, including particularly **copyright protection**. In this case, the question might reasonably be raised as to whether and to what extent the use of copyright protected material is lawful in the absence of any consent or licence from the copyright holder. Will an AI respect open data licences? Would it need to?

A comparable problem presents itself with respect to **fundamental rights** in general and the **right to data protection** in particular: when an AI is trained on data that contains personal data (i.e. information that can be linked to a specific natural person), is this lawful under European data protection legislation? What would be the legal basis and how can the principles of data protection law be observed when training the AI and when allowing it to be used?

Similarly, there are questions of **product liability and product quality**: who is ultimately responsible for ensuring that an AI is trustworthy and what does trustworthiness actually imply in general purpose AIs that have no explicitly defined usage limitations? How can risks of a particular AIS be identified and managed?

From the output side, the same topics can be examined from a different angle. Is an AI **capable of producing original works that are subject to intellectual property rights protections**, given that those new works are not created by a human being and that they are generated by introducing prompts to the AI, which will then try to recall and combine patterns from pre-existing works?

Equally importantly, how can the outputs of AIs be used in a manner that is fully respectful of the EU's **fundamental rights framework**, given that AIs can also be used in very sensitive contexts, such as healthcare (e.g. the identification of tumours) or public administration (e.g. the detection of fraud in relation to public resources)? Who is ultimately responsible in the event of failures, when it may be complex to determine whether the problem lay with training data, the AI algorithm itself, the context in which it was used or a lack of diligence in the individual user?

And what are the legal requirements for **bringing an AI product to market**, or for using it in a particular company or public administration?

There is thus a plethora of legal and policy questions for which **there is not always a clearly defined answer yet**. Part of the solution, as will be extensively discussed in this paper, may come from the EU's proposed AI Act, which was approved by the European Parliament on 13 March 2024. The act is still undergoing final checks and is expected to be adopted and published before the end of the current EU legislature.

The objective of this paper is to provide an overview of some of the main legal questions and currently available answers, building on a webinar series organised by the official portal for European data (data.europa.eu). The webinars focused on three topics in particular, which will also be examined in detail in this paper:

- data ownership, data use and legal insights in relation to intellectual property rights;
- fundamental rights, ethics and data protection;
- the regulatory approach of the EU in the emerging AI Act (AIA).

It goes without saying that neither the webinars nor this paper were exhaustive and other legal topics could still be examined in greater detail. The objective is, however, not comprehensiveness, but rather to obtain an accurate and representative overview of some of the main legal and policy challenges today.

This legal research paper is **intended as a resource for data policymakers, AI companies and the general public**.

- **Policymakers** can get a better understanding of the risks and opportunities in AI usage, and which legal risks and constraints to take into consideration.
- **AI companies** can get insights into the legal concerns and constraints surrounding AI, including their use of training data and requirements for bringing their products to market.
- **The general public** can learn how AI already affects them, and what their protection mechanisms are, under current and future law (such as the AIA).

3. Working definitions in this paper

This paper relies on a few important concepts that don't always have a clear or universally accepted meaning. To minimise misinterpretation, the following working definitions are used, which were based on the most current version of the proposed AIA.

Concept	Working definition
Artificial intelligence system	A machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments.
Training data	Data used for training an AIS through fitting its learnable parameters.
General purpose AI model	An AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are released on the market.
General purpose AI system	An AIS based on a general-purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration into other AISs.

Artificial intelligence and fundamental rights

1. Introduction

AISs have undeniably ushered in a new era, reshaping the fabric of our society with their transformative capabilities. From enhancing economic efficiencies through streamlined processes and reduced costs to enabling breakthroughs in research, facilitating autonomous transportation and powering smart home appliances, the breadth of opportunities presented by AI-based technologies is boundless. Indeed, these innovations stand as a beacon of hope, offering invaluable assistance in addressing some of the most pressing challenges of our time. However, amidst their promise lies a crucial caveat: the potential for significant, and sometimes catastrophic, impacts on both individual rights and societal well-being, if deployed without due consideration for fundamental human rights. With their ability to amass vast troves of personal data, AISs may have a significant impact on individual rights. These impacts encompass various areas of concerns, including personal autonomy, freedom of expression and the prevention of discrimination. Among the myriad impacts of AI, privacy and data protection emerge as the twin pillars most prone to being affected by AI's technological advancements.

As we delve deeper into the intricacies of personal data processing by AISs, it becomes increasingly imperative to establish a comprehensive understanding of the broader legal framework governing data protection within the European Union. This foundation is crucial for understanding the detailed complexities and potential risks involved when AI intersects with fundamental rights.

2. Fundamental rights framework

Fundamental rights represent a set of inherent and legally protected human entitlements essential for upholding dignity, equality and freedom. Within the European context, fundamental rights encompass a broad spectrum of civil, political, economic and social dimensions. These rights guarantee various aspects of human existence, including the right to life and integrity, liberty and security, privacy, freedom of expression and religion, education, non-discrimination and equality before the law. They serve as the bedrock of democratic societies, ensuring that individuals can live with autonomy and respect for their human dignity. The fundamental rights framework in Europe is underpinned by several key elements. At its core lies the Charter of Fundamental Rights of the European Union (the charter) which codifies the extensive array of rights and freedoms guaranteed to all individuals within the European Union. The charter, along with the European Convention on Human Rights, holds significant legal weight and serves as the primary source of fundamental rights law and policy within the EU. This framework additionally draws strength from international human rights instruments, such as the Universal Declaration of Human Rights (1948) and major UN human rights conventions, which provide further guidance and standards.

While AI can impinge upon various fundamental rights (such as individual personal autonomy or the right to be free from discrimination), the salience of its threats to privacy and personal data emerges notably due to AI's heavy reliance on data. In Section 3, we provide a general background on the legal framework for data protection in the European Union. A basic understanding of this framework is crucial for understanding the interplay between the application of AI and the fundamental right to privacy and the protection of personal data.

3. Privacy and data protection as a fundamental right in Europe

Throughout history, various civilisations have recognised the importance of personal privacy and data protection. Over centuries, societies have developed increasingly sophisticated understandings of privacy and data protection, reflecting evolving cultural norms and technological advancements. Ancient civilisations such as the Roman Empire had laws protecting the confidentiality of correspondence, emphasising the value of private communication. Similarly, the Magna Carta, signed in 1215, established principles of individual rights and liberties, laying the groundwork for modern concepts of privacy and data protection. During the Enlightenment period, thinkers such as John Locke and Jean-Jacques Rousseau emphasised the importance of individual autonomy and the right to privacy in their philosophical writings. These ideas influenced the drafting of modern legal frameworks, including the United States Constitution's Fourth Amendment, which protects against unreasonable searches and seizures. In the 20th century, the horrors of totalitarian regimes underscored the critical need for safeguards against government intrusion into personal lives, leading to the inclusion of privacy protections in international human rights instruments such as the Universal Declaration of Human Rights. These historical precedents demonstrate the enduring significance of privacy as a fundamental human right across different cultures and epochs. In the digital age, with the proliferation of data-driven technologies, concerns about privacy and data protection have become more pronounced, prompting legislative efforts worldwide to safeguard individuals' rights in an increasingly interconnected and data-centric world.

Throughout European history, personal data and privacy have been regarded as inherent rights, deeply ingrained in the fabric of society. These principles find expression in two complementary systems of fundamental rights protection: the Council of Europe's European Convention on Human Rights and the Charter of Fundamental Rights of the European Union and EU treaties.

Council of Europe legal framework

Although the right to privacy is not explicitly delineated as a standalone right within the European Convention on Human Rights (ECHR), its protection is enshrined in Article 8(1). This provision safeguards everyone's entitlement to respect for their private and family life, their home and their correspondence. Any governmental interference with these rights must be justified and proportionate. Given the expansive scope of personal data processing nowadays, it often intersects with an individual's right to privacy as articulated in Article 8(1) of the ECHR.

Additionally, the Council of Europe took a landmark step in 1981 by ratifying the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data, known as 'Convention

108'. This seminal agreement, updated in 2018, serves as a cornerstone of data protection in Europe. Convention 108 aims to uphold individuals' rights and fundamental freedoms, with a particular emphasis on the right to privacy, in the context of automated processing of personal data. By addressing the challenges posed by technological advancements, Convention 108 reinforces the Council of Europe's commitment to safeguarding privacy rights in an increasingly digitised world.

European Union legal framework

Within the European Union's legal framework, an extensive array of primary and secondary norms play a pivotal role in safeguarding personal information.

Among the primary norms, those enshrined in the charter hold particular importance. This foundational document allocates considerable attention to the subject matter, with two dedicated articles. Article 7 of the charter underscores the importance of respecting private and family life, along with the sanctity of communications and the home environment. Complementing this, Article 8 serves as a robust safeguard, offering explicit protection for personal data – a notable distinction from the ECHR, which lacks a dedicated article on data protection. It is noteworthy that Article 52(3) of the charter aims to establish coherence between the ECHR and the charter itself, specifying that when rights in the charter align with those protected by the ECHR, their interpretation and extent mirror those of the latter.

Expanding upon primary legislation, the European Union bolsters the protection of personal data through secondary legislation. The journey in EU data law began in 1995 with the adoption of Directive 95/46/EC, known as the data protection directive. This directive laid the groundwork for subsequent legislation, including Directive 2002/58/EC, commonly known as the e-privacy directive, which addresses personal data processing and privacy protection in the electronic communications sector. Notably, in 2016, the EU implemented the well-known general data protection regulation (GDPR), a landmark development in data protection law, which entered into full application as of May 2018. National laws of EU Member States (MSs) further complement this framework, ensuring that fundamental rights are upheld and respected at both European and domestic levels. Together, these elements form a comprehensive framework designed to protect the rights and dignity of individuals within the European Union and beyond.

General data protection regulation

The GDPR represents a significant milestone in data protection regulation, setting a new standard for privacy rights and accountability in the digital age. It stands as the most comprehensive and detailed framework to date, governing the collection, storage and processing of personal data. At its core, the GDPR establishes stringent obligations for entities that determine the purposes and means of data processing (data controllers) and for entities that provide services (processors), while simultaneously bestowing specific rights upon individuals, known as data subjects. By establishing clear rules and robust safeguards, it aims to foster trust and confidence in the handling of personal data, ultimately enhancing privacy and data protection for individuals within the European Union.

Prior to the GDPR, data protection laws within the European Union were fragmented and varied across MSs, resulting in inconsistencies and gaps in protection. The GDPR sought to harmonise these laws

and enhance privacy rights for individuals throughout the EU. Its overarching objective was to empower individuals to have greater control over their personal data and to ensure that organisations handling such data did so responsibly and transparently. One of the defining characteristics of the GDPR is its extraterritorial scope, which means that it applies not only to organisations operating within the EU but also to those outside the EU that process data of EU residents. This extended reach ensures that the protection of personal data is not confined by geographical boundaries, reflecting the global nature of data flows in the digital age.

Given the heavy reliance of AI on data, much of which may be personal in nature, it becomes imperative for developers and deployers of AISs to adhere strictly to the regulations and obligations stipulated by the GDPR. Personal data are useful at various stages of AIS development, including the training, testing and validation of AI models. During deployment, personal data can serve as input for predictions concerning individuals. Furthermore, outcomes produced by AISs may themselves qualify as personal data, as seen in scenarios like the derivation of an individual's risk score for developing a particular disease based on medical history, lifestyle patterns and genetic predispositions. Moreover, certain AI models may inherently consist of personal data, rendering such data indispensable for their effective functioning (e.g. in facial recognition systems, the AI model is built upon vast datasets containing images of individuals' faces; without access to such personal data, the AI model lacks the necessary foundation to perform its intended function effectively). Therefore, ensuring compliance with the GDPR becomes paramount not only in the handling of personal data used as input or generated as output, but also in the fundamental design and structure of AISs where personal data form an intrinsic part thereof.

The GDPR applies uniformly to all methods of processing personal data. However, the intricate operations inherent to AISs introduce unique complexities. While the GDPR provides a comprehensive framework for safeguarding personal data, the dynamic and evolving nature of AI development presents distinct challenges in upholding its principles. Therefore, a thorough examination of these principles is essential to understand the complexities and hurdles faced by AI developers in ensuring compliance within this rapidly evolving landscape.

This section delves deeper into the GDPR's fundamental principles and explains the difficulties of adhering to them within the dynamic and challenging realm of AI.

GDPR principles

The principles relating to the processing of personal data are enumerated and explained in Article 5 of the GDPR, and are explained below.

Lawfulness, fairness and transparency

The principle in general

Enshrined in Article 5(1), point (a), of the GDPR, this principle stipulates that the data processing must be lawful, fair and transparent to the data subject. Firstly, organisations must have a legitimate basis for processing personal data. Processing is lawful only when carried out under one or more of the legitimate grounds enumerated in Article 6(1) of the GDPR:

- the data subject's consent;
- the necessity to enter or perform a contract;
- the need to comply with a legal obligation;

- the protection of vital interests of the data subject;
- the performance of a task carried out in the public interest or the exercise of official authority;
- the legitimate interest of the controller or a third party.

Although all of the six items listed provide a valid legal ground for data processing, the two most frequently relied on are the first and the last ones (i.e. consent and legitimate interest).

The second fundamental aspect of the principle under consideration pertains to the obligation of controllers to transparently communicate to individuals the manner in which their data are used, commonly referred to as data processing. Controllers are mandated to uphold transparency and integrity in their dealings with data subjects, refraining from any form of misinformation or deception. They are entrusted with the responsibility to furnish data subjects with comprehensive details in accordance with Articles 12 and 13 of the GDPR. This entails disclosing the purpose of data processing, the duration of storage, the rights afforded to the data subjects, the categories of personal data involved, the origins of collected data if derived from external sources, the presence of automated decision-making processes, including profiling, alongside providing substantive insights into the underlying rationale, significance and anticipated implications for the individuals concerned. Such transparency not only fosters trust between controllers and data subjects but also ensures compliance with regulatory frameworks, thereby safeguarding individual privacy rights and promoting ethical data handling practices.

The principle in the context of artificial intelligence

In the context of AI development, adhering to the principle of lawfulness, fairness and transparency presents significant challenges. Firstly, the complexity of AI algorithms and their reliance on vast datasets make it difficult to ensure the legality and fairness of data processing activities.

AISs may inadvertently generate biased outcomes or make decisions based on incomplete or biased data, leading to unfair treatment of individuals. Additionally, the opacity of AI algorithms poses challenges to transparency, as understanding how AISs operate can be difficult (the 'black-box phenomenon'). In many instances, individuals may find themselves subjected to decisions made by AISs without a clear understanding of how or why those decisions were reached. This lack of transparency not only undermines accountability but also limits individuals' capacity to challenge or contest such decisions. The right to challenge decisions made by AISs is integral to safeguarding fundamental rights, yet it becomes increasingly elusive in the absence of transparent data processing practices.

Moreover, the lack of transparency in data processing can also impede an AI developer's ability to rely on certain legal grounds for processing. For instance, it may become challenging to obtain informed consent from data subjects when the processing activities within a given AIS are complex and the underlying logic is difficult to explain. In such scenarios, AI developers may be compelled to resort to alternative, albeit less certain, legal grounds for processing, such as legitimate interest. This underscores the complexity surrounding data processing in AISs and the importance of transparency in enabling individuals to make informed decisions about their data. This lack of transparency not only undermines trust but also obstructs individuals' capacity to exercise their rights under the GDPR, including the right to access and rectify their personal data (discussed further below).

Moreover, the swift advancement and widespread adoption of AI technologies often outstrip the development of regulatory frameworks, creating a formidable challenge for organisations to maintain compliance with evolving legal standards and uphold the fundamental principle of lawfulness. Reconciling the imperatives driving AI innovation with the imperative to safeguard data protection

principles demands continuous vigilance and proactive measures. Striking a delicate balance between technological progress and regulatory compliance necessitates concerted and sustained efforts to confront and resolve these inherent challenges.

Purpose limitation

The principle in general

One of the key tenets of the GDPR, the principle of purpose limitation embodied in Article 5(1), point (b), of the regulation, stipulates that personal data should only be collected for specified, explicit and legitimate purposes, and prohibits further processing of personal data for purposes that are incompatible with the purposes that led to the initial data collection.

By the same token, controllers shall refrain from collecting any personal data that are unnecessary, inadequate or irrelevant for these specified purposes. While subsequent processing for different purposes is not inherently prohibited, repurposing collected data is only permissible if the further processing aligns with the original purpose for which the data were initially collected (in which case no legal basis separate from that which allowed the initial collection of the personal data is required). For instance, Article 5(1), point (b), of the GDPR allows further processing of personal data for archival, historical research, or statistical purposes, presuming compatibility with the original purpose.

In order to assess whether the purpose of further processing is compatible with the purpose for which the personal data were initially collected, the controller should carry out a formal compatibility assessment of the intended further processing activity. This compatibility test should take into account several factors, such as:

- any link between the original purpose and the purpose of the intended further processing;
- the context in which the personal data have been collected, in particular the reasonable expectations of data subjects as to the further use of their data, based on their relationship with the controller;
- the nature of the personal data, in particular whether special categories of personal data are being processed;
- the consequences of the intended further processing for data subjects;
- the existence of appropriate safeguards in both the original and intended further processing operations.

Additionally, following a positive outcome of the compatibility assessment, the controller, prior to initiating the intended further processing, may be required to inform the data subject about the intended further processing activity, as the application of the principles set out in the GDPR (in particular the information of the data subject on those other purposes and on his or her rights, including the right to object) should be ensured.

There are two exceptions to this general prohibition on further processing for non-compatible purposes, namely, where further processing is based on data subject's consent or where further processing is based on an EU or MS law which the data controller is subject to. In these two cases, further processing is allowed under the GDPR, irrespective of the purpose compatibility (in other words the controller is presumed to be allowed to further process the personal data irrespective of the compatibility of the purposes).

The principle in the context of artificial intelligence

In the realm of AI, adhering to the principle of purpose limitation poses significant challenges for controllers. Defining the potential uses of collected data upfront is often exceedingly difficult, as processing purposes can remain ambiguous during the initial stages of data collection. Consequently, it has become commonplace in AI development to repurpose data at later stages. AI models, initially trained for specific purposes, often uncover unforeseen correlations within datasets, leading to a complete shift in their intended use. Thus, requiring AI developers to predetermine data collection purposes before processing begins could simply stifle innovation.

Data minimisation

The principle in general

Embodied in Article 5(1), point (c), of the GDPR, this principle seeks to restrict the indiscriminate collection of personal data. It mandates that only the minimal amount of personal data necessary for the intended purpose should be processed. Controllers are obliged to abstain from gathering data that are not directly and strictly relevant to the specified purpose or more than necessary.

The principle in the context of artificial intelligence

Observing the principle of data minimisation can be challenging for AI developers for several reasons. Firstly, this principle clashes with the very nature of AI-based technologies, which rely on the accumulation and analysis of massive amounts of data to function effectively. The basic functioning of AI models is grounded in their ability to learn from data, to draw inferences and to uncover correlations between various datasets. By definition, AI models require large datasets to effectively learn and generalise patterns. After all, the more data the AIS ingests, the more accurate its calculations and predictions will be. Additionally, the complexity and interconnectedness of AI algorithms may make it difficult to identify which specific data points are truly essential for achieving the desired outcomes. Consequently, developers of AISs may feel tempted to collect excessive amounts of data (including personal data) to enhance the accuracy of their AISs. Moreover, the lack of clear guidelines or standards for determining data relevance and necessity in AI development further complicates the adherence to the data minimisation principle.

Accuracy

The principle in general

Outlined in Article 5(1), point (d), of the GDPR, this principle implies the requirement that personal data be accurate and kept up to date at all times. Controllers are tasked with the responsibility of taking reasonable steps to ensure the accuracy of the data they process. They must regularly review personal data and promptly rectify or erase any inaccuracies, as processing inaccurate data may result in adverse consequences for the data subjects.

The principle in the context of artificial intelligence

Observing the principle of accuracy of personal data in the context of AI poses notable challenges for developers for many reasons. Firstly, AI algorithms often rely on vast and diverse datasets to train and refine their models, making it difficult to ensure the accuracy of every data point. AISs feed on data from various sources, however, the more diverse the sources, the higher the likelihood of encountering inaccuracies. Additionally, AISs may encounter issues with data quality, including errors, biases and inconsistencies, which can compromise the accuracy of the resulting insights and predictions. While some level of inaccuracy in the data used as input or the data produced as output of the AI models is accepted (as they aim to discover general tendencies or trends), such inaccuracies may harm individuals when they are used to create profiles or deliver inferences about those

individuals. Moreover, AI algorithms may uncover unexpected correlations or patterns in data that challenge conventional notions of accuracy, requiring careful interpretation and validation by human experts. Furthermore, the dynamic nature of data in AI applications, with continuous updates and changes, presents ongoing challenges in maintaining data accuracy over time. Last, but not least, given the prevalence of cyber threats, there is a significant risk of malicious actors targeting the AIS and tampering with the data used to train the AI model, potentially leading to inaccurate outputs.

Storage limitation

The principle in general

This principle, outlined in Article 5(1), point (e), of the GDPR, emphasises that personal data should only be retained in a manner that allows for the identification of data subjects for as long as necessary to fulfil the purposes for which the data was collected. Put simply, controllers must ensure that the duration of data retention aligns proportionately with the original objectives of data collection and is limited in time. Extending data storage beyond this period may be permissible solely for archiving purposes in the public interest, scientific or historical research purposes or statistical purposes, provided that appropriate safeguards are in place.

The principle in the context of artificial intelligence

Observing the principle of storage limitation, as outlined in the GDPR, presents notable challenges in the development of AI-based technologies. Firstly, as already noted, AISs often require vast amounts of data to train and refine their models, which inevitably leads to concerns about the storage of personal data beyond what is strictly necessary for the intended purposes. For example, AI-powered applications in healthcare may accumulate extensive patient data for predictive analytics, leading to questions about the retention period for historical medical records. The dynamic and iterative nature of AI development further complicates adherence to storage limitations, as the ongoing refinement of algorithms may necessitate the retention of historical data for continuous improvement. Also, collaborative research and development efforts in AI often involve data sharing among multiple stakeholders, resulting in the accumulation of extensive datasets across various platforms and organisations. This raises questions about the appropriate storage duration and scope, particularly in cross-border collaborations where differing regulatory requirements may apply. Furthermore, the potential for unintended data retention in AISs, such as cached or redundant data stored in memory or temporary storage, poses challenges in ensuring compliance with storage limitation requirements.

Integrity and confidentiality

The principle in general

Enshrined in Article 5(1), point (f), of the GDPR, this principle mandates that personal data must undergo processing in a manner that guarantees the security of the information. This entails safeguarding against unauthorised or unlawful disclosure or access to processed personal data (the confidentiality aspect) and protecting against accidental or unlawful alteration of or damage to personal data (the integrity aspect). Additionally, measures must be in place to prevent unintentional or unlawful loss of access to or destruction of personal data (the availability aspect). Among the most crucial techniques to ensure a high level of security are the encryption and pseudonymisation of personal data.

The principle in the context of artificial intelligence

Ensuring adherence to the principles of integrity and data confidentiality presents considerable hurdles during the development of AI technologies. Firstly, AISs frequently operate using extensive

datasets comprising sensitive personal data, heightening the vulnerability to unauthorised access, disclosure or alteration of information. For instance, AI applications used in the health sector or in the financial sector may handle confidential financial data, necessitating stringent measures to uphold the confidentiality and integrity of such data to prevent unauthorised access or tampering. Moreover, the interconnected nature of AISs, often reliant on shared data sources and collaborative training processes, further complicates the preservation of data integrity and confidentiality. Collaborative AI development initiatives, involving multiple stakeholders and data-sharing arrangements (such as cross-border research and innovation projects), may introduce weaknesses in data security and confidentiality, particularly when exchanging sensitive information across organisations and borders. Furthermore, the intricate nature of AI algorithms and their susceptibility to adversarial attacks amplify the challenges of safeguarding data integrity and confidentiality. These attacks, including techniques like data poisoning and model inversion, exploit AIS vulnerabilities to compromise data integrity or expose confidential information. Addressing these challenges demands robust implementation of data encryption, access controls and security protocols throughout the AI development life cycle, to ensure the integrity and confidentiality of data.

Accountability

The principle in general

Outlined in Article 5(2) of the GDPR, this principle mandates that controllers are responsible for demonstrating compliance with the GDPR's principles and for implementing appropriate measures to ensure compliance. These measures include in particular data protection impact assessments and maintaining detailed records of processing activities and will be addressed in more detail below.

The principle in the context of artificial intelligence

While Article 5(2) of the GDPR places the onus on controllers to demonstrate compliance with the GDPR's principles and to implement appropriate measures, to ensure adherence, the dynamic nature of AISs substantially complicates those accountability efforts. Firstly, the intricate algorithms and machine learning processes inherent in AISs often result in complex decision-making processes that are difficult to trace or explain. This opacity can hinder controllers' ability to fully understand and document the underlying mechanisms behind AI-driven decisions, thus impeding their ability to demonstrate compliance.

Additionally, the sheer volume and variety of data processed by AISs pose challenges in conducting comprehensive data protection impact assessments. AI algorithms may ingest vast amounts of data from diverse sources, making it challenging for controllers to assess the potential risks to individuals' privacy and ensure compliance with GDPR requirements. Moreover, the evolving nature of AI technologies introduces uncertainty regarding the adequacy of existing accountability measures. As AISs evolve and adapt over time, controllers must continuously reassess and update their compliance strategies to effectively mitigate risks and ensure accountability. Furthermore, the collaborative nature of AI development, involving multiple stakeholders and data-sharing agreements, further complicates accountability efforts. Ensuring accountability among various stakeholders and organisations in different countries involved in AI development requires robust governance structures and clear delineation of responsibilities.

As demonstrated above, upholding the fundamental principles of the GDPR may prove challenging in the realm of AI development. Addressing these challenges necessitates proactive efforts from the AI providers throughout the AI development life cycle. Only through such concerted efforts can

organisations effectively navigate the complexities of AI development while upholding the principles of data processing outlined in the GDPR.

Data subject's rights

In addition to the fundamental principles of data processing, the GDPR grants data subjects a range of rights to empower them in relation to their personal data. These rights, which can be invoked by data subjects, whose personal data is processed in the context of AI development and deployment, are discussed below.

Rights in relation to automated decision-making and profiling

The GDPR includes safeguards aimed at mitigating the risks associated with automated decision-making and profiling. These are especially meaningful in the context of AI, given how many decisions and actions nowadays are executed without human intervention and facilitated by AISs.

Article 22 of the GDPR explicitly grants individuals the right to not be subject to decisions made solely through automated processes, if such decisions have legal implications or similarly significantly affect them. This provision acknowledges the potential consequences of algorithmic decision-making on individuals' rights and seeks to ensure accountability and transparency in automated processes. With the increasing reliance on AISs to make critical decisions in various domains, such as finance, healthcare and employment, the protection afforded by this right becomes increasingly important. It underscores the need for AISs to operate ethically and transparently, with mechanisms in place that allow individuals to challenge automated decisions and understand the rationale behind them. Furthermore, the right not to be subject to automated decision-making underscores the importance of human oversight and accountability in the development and deployment of AI technologies. While AISs can offer efficiency and innovation, they must also respect individuals' rights and ensure fair and equitable treatment for all. Therefore, AI developers must implement robust mechanisms for oversight, accountability and transparency to uphold individuals' rights and prevent potential harms arising from automated decision-making and profiling.

Right to access

Individuals have the right to obtain confirmation as to whether or not their personal data is being processed and, if so, to access that data and information about how it is being processed.

In the context of AI, this right takes on added significance and complexity. Data subjects have the right to obtain confirmation from controllers as to whether their personal data is being processed and, if so, to access that data and relevant information about this processing. However, in the realm of AI, accessing personal data may not always be straightforward due to the intricate nature of AI algorithms and the vast amounts of data processed. AISs often operate on extensive datasets, with personal data used by them being dispersed across multiple platforms, databases or organisations. Consolidating and accessing this fragmented data can be complex, especially when data interoperability issues or data silos exist. Furthermore, controllers who are AI developers may face resource constraints or technical limitations when responding to data access requests. Processing large volumes of data to respond to data access requests may require significant time, resources and expertise. Therefore, ensuring effective access to personal data in the context of AI requires controllers to implement

transparent and user-friendly mechanisms that enable data subjects to understand and exercise their rights effectively.

Right to rectification

Linked to the controller's obligation to maintain accurate and up-to-date data, the right to rectification empowers data subjects to request the correction (i.e. rectification or completion) of inaccurate or incomplete personal data held by controllers. This right remains relevant across all stages of the AIS life cycle. For instance, during the development phase, data subjects can seek the correction of their information contained in the training dataset. Similarly, during the deployment phase, they may contest the accuracy of the outputs generated by the AISs.

The predictions and inferences generated by AISs often involve personal data, as defined in Article 4(1) of the GDPR. This includes both direct identifiers, such as names and addresses, and indirect identifiers or information that, when combined with other data, can identify an individual. However, rectifying the output of an AIS can be challenging as it primarily comprises statistical predictions rather than factual statements (even though the outputs may often be presented or interpreted as factual statements). Prediction scores are not inherently inaccurate merely because the factual reality doesn't match the prediction (e.g. a 99.5 % percent change of a cancer being present can be a reasonable and correct estimate, even if no cancer is detected afterwards); therefore, depending on the context and the presentation, the right to rectification may not apply if the personal data is not factually incorrect.

Right to erasure / right to be forgotten

The GDPR in Article 17 grants data subjects the right to request the erasure of their personal data under certain circumstances. When a data subject exercises this right, the controller is obligated not only to delete the data that they have processed directly but also to notify all other known recipients with whom they shared the data about the data subject's request. This right can only be exercised in certain limited instances, for example, when the data is no longer necessary for the purposes for which it was collected or if the processing is unlawful. It can also be exercised by data subjects who object to the processing of their data and for whom the controller cannot demonstrate other overriding legitimate grounds for further processing.

Exercising this right within the realm of AI might be a tough nut to crack. AISs often incorporate vast amounts of data from diverse sources located in various locations. Data is usually replicated across multiple systems for backups. All of this makes it difficult to track and identify specific instances of personal data for erasure. Moreover, the dynamic and evolving nature of AI algorithms complicates the erasure process, as data may be continuously processed and integrated into AI models over time. The source data can become increasingly difficult or even impossible to find or remove. In order to be able to entirely erase one's personal data included in an AI model, it may be necessary to retrain the AI model based on a data set that no longer includes the erased data and is not influenced by the 'algorithmic shadow' of that individual's data.

This, however, might not be feasible due to the substantial computational and engineering expenses, along with time limitations, particularly concerning complex AI models. Additionally, the inherent opacity of AI decision-making processes may hinder data subjects' ability (or indeed any party's ability) to determine whether their personal data has been completely erased from AISs. The proliferation of AI-based applications across various sectors and industries also raises concerns about the widespread dissemination and potential replication of personal data, further complicating the erasure process. Exercising the right to erasure may also be problematic, due to uncertainties regarding the scope of the request.

Specifically, it may be unclear whether the request should only pertain to the data directly provided by the data subject or also encompass the data derived or inferred from that initial dataset. This ambiguity raises questions about the extent to which AISs should erase not only the raw data but also any insights, predictions or conclusions drawn from it. The reference case on this right in the EU is the C-131/12 case, commonly known as the Google Spain case. In this landmark ruling, the claimant requested the removal of certain search engine results generated by Google's algorithm. These results were based on inferences drawn from the claimant's personal data. The Court of Justice of the European Union (CJEU) ruled in favour of the claimant, affirming the individual's right to have such derived data erased from the search engine (but not from the original websites where the data were hosted). This case underscores the significance of ensuring that data erasure requests extend beyond just the raw data to encompass any derived or inferred information generated by AI algorithms.

To address the above challenges, controllers should seek to design their AISs in a way that the deletion requests can be effectuated, in accordance with the principle of privacy by design. They should implement robust data governance practices and transparency mechanisms to ensure the effective erasure of personal data from AISs. Additionally, clear guidelines and standards should be established for the secure and permanent deletion of personal data within the context of AI development and deployment.

Right to restriction of processing

A substitute to the right to erasure, the right to restriction of processing, grants individuals the authority to limit the processing of their personal data under specific circumstances, such as when the accuracy of the data is contested or when the processing is unlawful. As a result, controllers must limit the processing operations they carry out on the data and may only store it.

The concerns with exercising the right to restriction of processing are similar to those related to the right to erasure. Due to the fact that AISs operate on extensive datasets sourced from diverse channels, it may be especially intricate for individuals to pinpoint and control the processing of their specific personal data. Also, the dynamic nature of AI algorithms, continually learning and evolving from new data inputs, complicates efforts to enforce processing restrictions effectively. The aforementioned opacity inherent in AI decision-making processes exacerbates the challenge, as individuals may struggle to monitor and enforce limitations on the processing of their personal data by AISs. Additionally, the interconnectedness of AISs across various platforms and networks may lead to inadvertent processing of restricted personal data beyond the intended scope.

To address these challenges, there is a pressing need for enhanced transparency and communication mechanisms to empower individuals in monitoring and enforcing restrictions on their personal data processed by AISs. Furthermore, controllers must establish robust controls and mechanisms within AISs to facilitate data subjects' in exercising their right to restrict processing effectively, ensuring compliance with data protection regulations and upholding individuals' privacy rights.

Right to data portability

The right to data portability, enshrined in Article 20 of the GDPR, enables individuals to obtain their personal data in a structured, commonly used and machine-readable format and to transfer that data between different services or platforms.

In the AI setting, exercising this right might present some significant hurdles. Firstly, personal data derived from further examination of provided information is exempt from the right to portability. This signifies that the outcomes generated by AI models, such as predictions and classifications regarding

individuals, lie outside the purview of portability rights. In certain instances, some or all of the characteristics used to train the model may have originated from prior analysis of personal data.

For example, a credit score obtained through statistical analysis based on an individual's financial data might subsequently be employed as a feature in a machine learning model. In such cases, the credit score is not encompassed within the scope of data portability rights, even if other attributes are. Secondly, extracting and transferring personal data in a usable format from complex and interconnected datasets may be particularly challenging. At the same time, the proprietary algorithms and formats used by AISs may not be readily compatible with other services or platforms, hindering seamless data portability. Moreover, the dynamic nature of AI algorithms, which continuously evolve based on new data inputs, adds an additional layer of complexity to the portability process. Individuals may struggle to ensure the accuracy and completeness of their transferred data, particularly when dealing with AI-driven insights and predictions that are constantly evolving.

Overcoming these challenges requires the development of standardised data formats and interoperability protocols tailored to AISs, along with enhanced transparency and accountability mechanisms to facilitate individuals' in exercising their right to data portability effectively.

Right to object

A fundamental provision of the GDPR, which empowers individuals to object to the processing of their personal data on grounds related to their particular situation, when the processing is necessary for the performance of a task carried out in the public interest or in the exercise of official authority vested in the controller or when the processing is necessary for the purpose of legitimate interest of the controller or a third party. Moreover, individuals can object to the processing of their personal data for marketing purposes.

Exercising this right leads to unique obstacles in the context of AI. Unlike traditional data processing methods, AISs often operate autonomously, with their internal decision-making processes being opaque or difficult to interpret, commonly referred to as 'black boxes'. These systems rely on complex algorithms and extensive datasets to make decisions, which may lead individuals to struggle in comprehending the logic behind AI-driven decisions and identifying instances where their data is being processed in ways they object to. AISs might produce conclusions and forecasts derived from intricate connections and patterns in the data, complicating individuals' ability to determine if their objections are justified or relevant. When decisions are made autonomously by AISs, individuals may find it challenging to identify who to address their objections to and how to effectively communicate their concerns. Moreover, the widespread adoption of AI across diverse applications and sectors further complicates exercising the right to object. Individuals may interact with multiple AISs operated by different entities, making it difficult to centrally manage objections and ensure consistent compliance with data protection preferences.

Addressing these challenges requires enhanced transparency, accountability and accessibility measures to empower individuals to assert their rights effectively in the AI-driven digital landscape.

Other accountability mechanisms

The GDPR not only grants extensive rights to data subjects but also establishes a robust framework of control aimed at safeguarding these rights. Central to this framework are the accountability and oversight obligations imposed on controllers. These accountability obligations not only enhance

transparency and trust but also reinforce the protection of individuals' rights. They serve as a cornerstone of the GDPR's regulatory approach, ensuring that controllers are held responsible for their data processing activities and that appropriate measures are in place to protect individuals' personal data. Below, we examine those other accountability obligations that the GDPR imposes on controllers, with a specific focus on those controllers who are AI developers.

Data protection impact assessment

As stipulated in Article 35 of the GDPR, controllers are obligated to conduct data protection impact assessments (DPIAs) for data processing activities that are likely to result in high risk to individuals' rights and freedoms. DPIAs are systematic assessments aimed at identifying, assessing and mitigating the risks associated with data processing. They are particularly important when implementing new technologies or processing sensitive personal data. The purpose of DPIAs is to ensure that controllers proactively address privacy risks and comply with data protection principles. They involve assessing the necessity and proportionality of data processing activities, evaluating potential risks to data subjects and implementing measures to mitigate identified risks. Given the complexity and potential implications of AI technologies, DPIAs are particularly crucial in this setting. AISs depend on massive amounts of data and complex algorithms to make decisions or predictions, which can pose significant risks to individuals' privacy and rights.

Controllers who are AI developers must therefore carefully assess the potential risks associated with AI-driven data processing activities, including the potential for bias, discrimination, or infringement of individuals' rights. DPIAs in the AI sector involve evaluating the transparency and fairness of AI algorithms, assessing the potential impacts on individuals' rights and implementing measures to mitigate identified risks. By conducting DPIAs, AI controllers demonstrate their commitment to accountability and transparency in AI development and deployment, ensuring that individuals' rights are adequately protected in the rapidly evolving landscape of AI technologies.

The upcoming AIA introduces a new requirement under Article 29a, mandating deployers of high-risk AISs (HRAIS) to conduct a fundamental rights impact assessment (FRIA) to evaluate the potential impact of AISs on fundamental rights. Unlike a DPIA under the GDPR, which focuses primarily on data protection risks, a FRIA considers broader societal implications, including ethical, social and fundamental rights considerations, ensuring a more comprehensive evaluation of AI deployments, and should be conducted in conjunction with a DPIA.

Record of processing activities

Article 30 of the GDPR outlines the requirements for controllers to maintain comprehensive records of processing activities. These records serve as a vital repository of information, encompassing key details essential for ensuring compliance with data protection regulations. The records must include details such as the contact information of the controller, joint controller, representative, and data protection officer, where applicable. Additionally, they should delineate the specific purposes behind the data processing activities and provide a thorough description of the categories of data subjects and personal data involved. Furthermore, the records must document the categories of recipients to whom the personal data have been or will be disclosed, including any transfers to non-EU countries or international organisations, along with the requisite safeguards. Anticipated timeframes for the erasure of different data categories should be outlined whenever feasible, alongside a general overview of the technical and organisational security measures implemented to safeguard the data.

Maintaining a comprehensive record of processing activities poses significant challenges for AI developers, mainly due to the intricate and multifaceted nature of processing operations within AISs.

The complex algorithms and iterative nature of AI development make it arduous to accurately document all data processing activities, particularly with the vast array of data sources and evolving models involved. Furthermore, the decentralised structure of AI development teams and the involvement of numerous stakeholders add layers of complexity to the task, making it even more challenging to maintain thorough records of processing activities.

Data Protection Officer

As stipulated in Article 37 of the GDPR, some controllers and processors are also obliged to designate a data protection officer (DPO). The DPO plays a crucial role in overseeing the organisation's data protection strategy and ensuring compliance with data protection laws. Responsibilities include advising on data protection obligations, monitoring compliance, guiding data protection impact assessments, and acting as a contact point for data subjects and supervisory authorities. Mandatory appointment of a DPO applies to public authorities, organisations engaged in systematic monitoring of data subjects on a large scale, and those processing special categories of personal data extensively.

Given that AISs often manage exorbitant datasets and may involve systematic monitoring or processing of sensitive personal information, AI providers may be subject to the DPO requirement under the GDPR. However, AI developers may face challenges in appointing a DPO, as finding an individual with both expertise in data protection regulations and a deep understanding of the complex nature of AI can be particularly daunting.

4. Example case: navigating the impact of artificial intelligence in healthcare

Introduction

Building upon the foundational principles of the GDPR in the context of AI, this chapter delves into the implications of AI with a focus on the healthcare sector. The choice to spotlight the healthcare industry stems from its unique position as both a pioneer and a significant beneficiary of AI technologies. Healthcare represents a domain where the integration of AI has rapidly evolved in the recent years, transforming traditional practices and opening new avenues for improved patient care, diagnosis and treatment. Furthermore, healthcare data is inherently sensitive and highly regulated, making it a prime example to elucidate the complex interplay between AI advancements and privacy concerns. With the increasing digitisation of medical records, the proliferation of wearable health devices and the adoption of telemedicine platforms, the healthcare sector offers a rich landscape to examine AI's impact on privacy rights. This chapter explores the challenges and opportunities of AI integration in healthcare, providing insights into maintaining privacy safeguards amidst rapid technological advancements. It examines the reasons for the rapid adoption of AI in healthcare, offers real-world examples of successful AI applications, discusses privacy and security risks and proposes mitigating measures to protect individuals' privacy rights.

Artificial intelligence in healthcare: a short landscape overview

The adoption of AI in healthcare has surged in recent years, driven by several factors. AI tools have the capacity to enhance accuracy, minimise expenses and streamline processes in contrast to conventional diagnostic methods. Furthermore, AI has the potential to mitigate the likelihood of human errors while delivering more precise outcomes within shorter timeframes. It offers unparalleled capabilities to process and analyse large volumes of healthcare data, including electronic health records (EHRs), medical images and genomic data, at speeds and scales beyond human capacity. This ability enables healthcare providers to extract valuable insights from complex datasets, leading to more accurate diagnoses, personalised treatment plans and improved patient outcomes. This rise in digital health technologies, coupled with the increasing demand for remote healthcare services, has accelerated the integration of AI-driven solutions into clinical practice. Telemedicine platforms, wearable devices and mobile health applications leverage AI algorithms to deliver virtual consultations, remote monitoring and predictive analytics, enhancing access to healthcare services and empowering patients to take control of their health.

Numerous examples of EU funded research and innovation projects can illustrate the application of AI in various healthcare settings.

- Oncorelief is a Horizon 2020 project which developed a user-centred AIS designed to function as an intuitive smart digital assistant, which aims to revolutionise post-treatment care by providing personalised support tailored to each cancer survivors' needs. This AI-driven assistant not only assists with post-treatment activities and tasks but also proactively suggests actions to improve the patients' overall health, wellbeing and active healthcare engagement. By facilitating a continuous wellness journey, the Oncorelief assistant ensures that cancer survivors remain actively involved in maintaining their health during the critical post-treatment period, promising to enhance long-term health outcomes and quality of life for survivors.

- Rebecca is a Horizon 2020 project which aims to leverage real-world data to enhance clinical research and improve current clinical practices. By integrating clinical data with information on patients' daily behaviours like physical activity, diet, sleep and online interactions collected through mobile and wearable devices, Rebecca generates new insights. It creates novel functional and emotional indicators for each patient to assess their well-being and quality of life, thereby optimising their care. The Rebecca 360° platform, comprising unobtrusive mobile applications, supports breast cancer survivors in their daily lives and facilitates their communication with healthcare professionals. It also contains information on future post-cancer treatment guidelines and practices.

- Oncoscreen, a Horizon Europe research and innovation project, is dedicated to developing AI-driven solutions for personalised colorectal cancer screening and early, non-invasive and cost-efficient detection. By integrating advanced AI algorithms with cutting-edge medical diagnostic and imaging technologies, Oncoscreen aims to revolutionise colorectal cancer diagnosis, enabling early intervention and improved patient outcomes.

- LUCIA is another EU-funded project which aims to understand and discover new risk factors that contribute to the development of lung cancer; AI models are used to identify environmental, biological, demographic, community and individual level risk factors associated with the formation of lung cancer, by combining open data sources (e.g. environmental data) with retrospective clinical data from clinical partners, and prospective data collected during clinical studies, data collected via medical devices and through patient

questionnaires. Additionally, the AI models help determine risk scores for lung cancer, which can be used to screen patients and detect lung cancer at an early stage.

Risks associated with the use of artificial intelligence in healthcare

Despite its transformative potential, the widespread adoption of AI in healthcare raises significant privacy and security concerns. In fact, security and patient privacy are the core concerns in the healthcare sector when it comes to AI, as access to patient medical data is central to the training of AI models and the use of AI-based solutions in the delivery of healthcare. The growing prevalence of AI solutions and technology in healthcare, highlighted most recently by the COVID-19 pandemic, has demonstrated the potential for significant ramifications on the rights of patients and citizens.

One of the primary risks is the risk of personal data being shared and used without the patient's explicit consent. As stipulated in Article 9 of the GDPR, processing of personal data concerning health shall be by default prohibited. Such processing shall only be allowed under certain conditions enumerated in Article 9(2) of the GDPR. The most commonly invoked condition among these, is the case where the data subject has given explicit consent for the processing of those personal data for one or more specified purposes. In reality, however, AISs often analyse and process personal health information without individuals' informed consent.

Another persistent concern is data repurposing, also known as 'function creep'. This phenomenon involves the unauthorised or unintended use of data collected for one purpose being repurposed for other unrelated or unexpected ends. A striking example of function creep occurred in Singapore, where data collected through the government's COVID-19 tracing app, intended for public health monitoring and contact tracing, was repurposed for unrelated endeavours, such as criminal investigations. Similarly, in Germany, COVID tracking data was used by police to identify individuals who were present at a restaurant where a death occurred, demonstrating a concerning trend of expanding the use of collected data beyond its original purpose.

Furthermore, the integration of AI-driven technologies and the reliance on them in healthcare introduces cybersecurity risks, encompassing cyberattacks targeting AISs, data breaches leading to identity theft or medical fraud and the exploitation of AI algorithm vulnerabilities to manipulate medical decisions or endanger patient safety. An illustrative case is the September 2020 cyberattack on Dusseldorf University Hospital, which interfered with the hospital's data and rendered the system inoperable. As a result, a patient could not be admitted to the hospital and had to be redirected to another facility in a distant city, which ultimately resulted in her demise. Although it was later argued that it could not be proven that the death was directly caused by the cyberattack, because the patient was already suffering a life-threatening condition, this case brought to the forefront the real physical harms that cyberattacks can cause in the healthcare sphere. Similarly, the Elekta case in April 2021 demonstrated how cyberattacks on AISs can directly impact patient rights, with a ransomware attack affecting 170 health systems in the United States (US) and delaying cancer treatment care nationwide. Additionally, AI-controlled personal medical devices, such as insulin pumps for diabetes patients, face hacking risks, potentially allowing remote manipulation and the administration of excessive insulin doses.

5. Risk mitigating measures: general strategies and approaches

To effectively mitigate the privacy and security risks associated with the deployment of AI in healthcare, a multifaceted approach is essential. Firstly, robust data protection mechanisms, such as encryption, pseudonymisation and access controls, should be employed to safeguard patient data against unauthorised access and breaches. Organisations must ensure awareness and understanding of data privacy and security risks, emphasising that AI developers and deployers should comply with applicable laws, such as the GDPR. Custodians of data must give top priority to safeguarding and discouraging alternative data usage in order to uphold the privacy and confidentiality of patients. Transparent and accountable AI governance frameworks should be established to ensure that AISs adhere to ethical principles, regulatory requirements, and best practices for data privacy and security.

Requiring organisations deploying AI to conduct FRIAs, as mandated for HRAISs by the pending AIA, while also conducting comprehensive data protection impact assessments (DPIAs) to identify and mitigate potential privacy risks associated with AI deployment, can further enhance privacy protection. Advocating for the use of synthetic data, artificially generated and disconnected from real individuals, could also enhance privacy and security by minimising the risks associated with real patient data. Ongoing research efforts to enhance AIS security and protect algorithms against cyberattacks are imperative. Continuous monitoring, auditing and evaluation of AISs' performance and compliance with privacy regulations are essential to proactively detect and mitigate any privacy breaches or security incidents.

Moreover, continuous staff training and awareness programs should be implemented to educate healthcare professionals about the importance of privacy protection and security measures when using AI technologies. Collaborative efforts between healthcare institutions, technology providers, regulators and policymakers are also crucial to establish standardised protocols, guidelines and regulations for the responsible development and deployment of AI applications in healthcare while ensuring the protection of patient privacy and data security. Through these concerted efforts, the healthcare industry can navigate the complexities of AI deployment while prioritising patient privacy and data security.

6. Conclusion

In this exploration of fundamental rights and data protection in the context of AI, we have delved into the intricate dynamics shaping the intersection of technology and human rights. It is clear that, while AI holds tremendous promise, its widespread adoption must be accompanied by robust privacy protections and regulatory safeguards. By adhering to GDPR principles, implementing privacy-enhancing technologies and adopting transparent and accountable AI governance frameworks, stakeholders can harness the transformative potential of AI while safeguarding individuals' fundamental right to privacy and data protection. As the AI landscape continues to evolve, it is imperative to find new and more effective ways strike a balance between innovation and privacy protection, and to ensure that AI-driven advancements benefit society, while respecting individuals' privacy rights.

Artificial intelligence and intellectual property – the (lack of) creativity of artificial intelligence, and its dependence on pre-existing inputs

The uptake of generative AI is reshaping our perception of creativity. As generative AISs continue to evolve, they are increasingly capable of producing outputs that blur the lines between human and machine-generated content. From generating texts, to creating art and music, generative AISs demonstrate remarkable potential in redefining traditional notions of creativity. However, this technological advancement also raises questions about copyright on both the input and output side. This chapter delves into the complex interplay between AI and copyright, and explores the potential impact of AI-generated output on the creative ecosystem. It aims to explain how copyright (and related rights) currently apply to generative AI, and to inspire future dialogue on the interaction between copyright and AI.

1. Artificial intelligence and training data: addressing copyright challenges

In the realm of AI, the importance of training data cannot be overstated. These vast datasets serve as the foundation upon which AISs learn, adapt and make decisions. However, the use of such datasets for training purposes raises questions regarding copyright and ownership. This is particularly the case for generative AISs, which are designed to learn patterns and structures from large datasets, and on the basis thereof, generate new data or content that mimics or resembles human-created content.

The foundation models of such generative AISs, including large language models (LLMs) and text-to-image models are often trained on datasets that include publicly available materials, such as web pages, images, articles, blog posts and tweets. Many of these materials are, however, not owned by the generative AIS's trainer and are potentially protected by copyright.

So what does this copyright protection entail? From a policy perspective, copyright is meant to encourage the creation of original works by providing the authors of such works with exclusive rights to control the exploitation of their work and protect its integrity. Encouraging the creation of original works contributes to the cultural, social and economic advancement of society and is therefore desirable.

A 'work' that is eligible for copyright protection can take various forms. Article 2 of the Berne Convention provides a list of literary and artistic works that are generally copyrightable. These include

books, dramatic works, musical compositions, choreographies, sculptures and cinematographic works. It is a common misconception however that copyright is only a matter for writers, composers and other artists. Essentially, copyright protects any work that is both expressed in a concrete form and original.

The first criterion entails that copyright protection may be granted to expressions, but not to mere ideas, procedures, methods of operation or mathematical concepts (even if they were original). Put differently, as has been affirmed by the European Court of Justice, in order for copyright protection to apply, a work must be expressed in a manner which makes it identifiable with sufficient precision and objectivity. Consequently, there will be no copyright infringement when you copy someone's idea or use it as an inspiration, as long as you give a different expression to this idea. Likewise, copyright protection is not granted to a 'style', 'genre', 'trend' or 'technique'. Making a work of anti-authoritarian street art for example, does not by definition imply an infringement of Banksy's copyright.

The second criterion of 'originality' entails that a work (or parts of a work) can only be protected by copyright if it (or they) contain(s) 'elements which are the expression of the intellectual creation of the author of the work'. This means that for copyright protection to exist, a work needs to be the author's own intellectual creation, reflecting their personality. As such, a work will only be copyright protected if the author has been able to express their creative abilities in the production of the work by making free and creative choices that stamp the work created with their 'personal touch'.

This means that for the purpose of copyright protection, it is entirely irrelevant if a work is pretty or ugly, if it has artistic value or not or what the quality of the work is. If a work is expressed in an original form, this automatically triggers copyright protection. If you yourself, for example, take a photo of almost anything, and you have made original choices regarding the perspective, shadow play, composition, colours, etc. in such a way that the photo expresses your own intellectual creation, then this photo will be copyright protected. Needless to say, the originality standard in EU copyright law is considered to be rather low.

Copyright protection in principle vests in the author or authors of the protected work, i.e. the person(s) whose intellectual creation the work(s) express(es). This means that copyright originally always vests in one or more physical person(s) and not in a legal person. However, this does not exclude that legal persons (such as companies) can hold copyright in works. Legal persons can become copyright holders through copyright transfer agreements (or exceptionally through specific legislative provisions). Copyright protection moreover persists for quite a long time. More precisely, a copyright holder can exercise its rights throughout the life of the author and for 70 years after his death. As a result, many works that are publicly available online, such as books, news articles, photos, videos, music and paintings, may (still) be protected by copyright. If such works are scraped from the Internet (i.e. found and locally stored using a fully automated tool), and further used as training data for generative AISs, this has copyright implications.

Indeed, around the globe, many artists have already expressed their strong dissatisfaction with generative AISs using their works for training purposes without authorisation, arguing that AIS providers are 'stealing' their intellectual property and are potentially even undermining their jobs. That is why some artists have organised themselves and founded a European Guild for AI Regulation to 'bring to the public attention how their data and intellectual properties are being exploited without

their consent, on a scale never seen before'. Moreover, several lawsuits for copyright infringement have been filed by copyright holders against AIS providers for allegedly scraping copyright protected works from the internet and using them in the creation of AI products.

This raises the question as to whether authors can control the use of their copyrighted works for training generative AISs through their copyright. Under EU copyright law (specifically the information society directive), the exclusive rights of a copyright holder consist of economic rights on the one hand and moral rights on the other. The economic rights include:

1. the exclusive right to authorise or prohibit the direct or indirect reproduction of a work by any means and in any form in whole or in part ('the reproduction right'); and
2. the exclusive right to authorise or prohibit any communication to the public of a work ('the right of communication to the public').

The moral rights are not fully harmonised at the EU level, but will in most MSs at a minimum include the right of the author(s):

1. to be identified as the author(s) of any work he/they create ('right of paternity'); and
2. to prevent others from subjecting their works to derogatory treatment, in each case for the duration of the works' copyright ('right of integrity').

Whereas the economic rights are transferable from the author to a third party, moral rights are understood to be so closely connected to the person of the author that only he can exercise them. As a result, moral rights are non-transferable.

When training a generative AIS, various techniques are used, including text and data mining (TDM). TDM refers to the automated processing of large volumes of text and data to uncover new knowledge or insights. TDM usually requires the copying of large quantities of material (that may be copyright protected), extracting the relevant data and recombining it to identify patterns. This is where the right to reproduction comes into play. Given that under EU copyright law the right to reproduction is interpreted in a very broad way, many of the copies that are made in the process of TDM are generally considered to be an 'act of reproduction' that falls under the reproduction right. As a consequence, all right holders of copyright protected works that are included in a training dataset in principle have to authorise the use of their works for the training of a generative AIS. In the absence thereof, copyright infringement would occur.

However, the exclusive rights of the copyright holder are not absolute. As the requirement to receive authorisation from a right holder may in some cases be overly burdensome for the user of a work (or even violate fundamental rights), under EU copyright law, a limited number of exceptions to the exclusive rights exist. These exceptions are meant to strike a balance between the exclusive rights of the right holder and safeguarding the public interest and users' fundamental rights. The exceptions, for example, permit the reproduction of works for private use, criticism or review, illustration for teaching or scientific research, and parody.

In the context of generative AISs, originally, two exceptions were of particular relevance, namely the exception for temporary reproductions, as found in Article 5(1) of the information society directive, and the exception for teaching and scientific research, as found in Article 5(3), point (a), of the information society directive.

The exception for temporary reproductions is a mandatory exception in all MSs. It essentially interprets the right to reproduction in a way that is compatible with our modern digital society. According to this exception, temporary acts of reproduction, which are transient or incidental and an integral and essential part of a technological process and whose sole purpose is to enable a transmission in a network between third parties by an intermediary or the lawful use of a work or other subject matter to be made, and which have no independent economic significance, may be undertaken without infringing copyright. Much like the copies that you make in your memory when reading a book do not require the authorisation of the copyright holder of the book, the exception for temporary reproductions ensures that also the temporary reproductions that are made in the memory and on the screen of a computer when browsing a website do not require the authorisation of the copyright holders of any works that are included in that website.

The exception for teaching and scientific research is an optional exception. It allows the use of copyrighted works for the sole purpose of illustration for teaching or scientific research without the authorisation of the right holder, as long as the source, including the author's name, is indicated, unless this turns out to be impossible and to the extent justified by the non-commercial purpose to be achieved. Like all other exceptions and limitations, the application of these two exceptions is subject to the 'three-step-test', which means that they can only be applied in certain special cases which do not conflict with a normal exploitation of the work or other subject-matter and do not unreasonably prejudice the legitimate interests of the right holder.

Given the many conditions to be fulfilled for these exceptions to apply, for some time, it was unclear if they could be relied upon in the context of reproductions for TDM purposes. Given this legal uncertainty concerning TDM, in 2019 the EU legislator adopted two new mandatory exceptions for TDM in Articles 3 and 4 of its 2019 directive on copyright and related rights in the digital single market (DSM).

Article 4 of the DSM directive provides for a general exception which allows anyone to perform TDM on copyright protected works that are lawfully accessible. Both public and private entities can benefit from the exception, and even TDM for commercial purposes is covered by the exception. Although the exception may thus seem to be very broad, it is subject to strict conditions that severely restrict its scope. First of all, reproductions made for the purposes of TDM may only be retained for as long as is necessary for those purposes. Secondly, the exception only applies on the condition that the use of works for TDM purposes has not been expressly reserved (read: prohibited) by the right holders in an appropriate manner, such as machine-readable means in the case of content made publicly available online. For works that are not publicly available online, a contractual agreement or unilateral declaration may also be an appropriate manner to reserve rights. This 'opt-out mechanism' obviously implies a significant weakening of the general exception, as text and data mining will not be allowed if right holders have prohibited the use for text and data mining purposes of their works, for example via a metatag, terms and conditions of a website or service, or a robot protocol (robot.txt) if the works are publicly available online.

Article 3 of the DSM directive provides for a specific TDM-exception for scientific research. Contrary to the general exception of Article 4, only research organisations and cultural heritage institutions can benefit from the exception, which allows them to make reproductions of works to which they have lawful access in order to carry out text and data mining for the purposes of scientific research. The

exception is mandatory, meaning that right holders cannot oppose text and data mining carried out by research organisations and cultural heritage institutions in the context of scientific research via an express reservation of rights. Nevertheless, right holders are allowed to apply measures to ensure the security and integrity of the networks and databases where their works are hosted, to the extent that these do not go beyond what is necessary to achieve that objective. Thus, although such measures should not interfere with the application of the exception, it is not unlikely that, in practice, measures to ensure the security and integrity of networks and databases may prevent research organisations from performing text and data-mining activities on works to which they have lawful access. In any case, beneficiaries of the exception may only store the copies of works that are made through text and data mining with an appropriate level of security for the purposes of scientific research, including for the verification of research results. In this context, the DSM directive also requires MSs to encourage right holders, research organisations and cultural heritage institutions to define commonly agreed best practices concerning secure storage and measures for the security and integrity of networks and databases that right holders may apply.

With these two new exceptions, the EU legislators have thus attempted to resolve the tensions between tech companies and authors regarding the use of their copyright protected works for AI training purposes. Whether this has been a success remains to be seen. Not only is it conceivable that these new exceptions will lead to difficulties in interpretation (for example, because the modalities of the 'opt-out' have not (yet) been standardised, and many national variations to the exceptions may arise), but it is also yet to be seen if authors and tech companies will find a compromise on how these exceptions are applied in practice.

Moreover, the exceptions have been scrutinised for putting the European Union's AI sector at a competitive disadvantage given that, in other jurisdictions, commercial TDM of copyrighted works may be allowed, for example, under the 'fair use' doctrine in the US, without the obstacle of an opt-out mechanism or other restrictive conditions. It is true that, as a result of the EU's TDM exceptions, EU-based AIS providers may in the future be confronted with authors that have opted-out of the TDM exception and will then be left with two options: incur additional licensing costs to be able to train their systems on the copyright protected works, or exclude the works from their training datasets altogether. It thus remains to be seen whether and how these differences between US and EU copyright law will impact future AI development.

2. Can an artificial intelligence be a creator? Dealing with (non)creative outputs

The section above examined the important questions that arise when AIs try to use copyright protected works as a training input. The intellectual property rights issue can also be examined from the other side: can the outputs of generative AISs be protected by copyright or related rights?

Many generative AISs today produce creative works that are hardly distinguishable from works created by humans. This prompts the question: are such creative machine-made productions also eligible for copyright protection? As explained above, in order for a work to be copyright protected, it

needs to constitute a concrete and original expression of an author's own intellectual creation. The author is the human who makes the free and creative choices for a work and expresses his personality in it. This inherently human-centric approach to EU copyright law entails that AISs currently cannot be authors of copyright protected works. Indeed, AISs as machines are not able to make creative choices that bring the output they create in the realm of copyright protection. As a result, works that are produced solely by AI ('AI-generated output') are not protected by copyright.

This does not, however, mean that the output of generative AISs is never protected by copyright. Where works are produced with an AIS, the relevant question is if the work is a (human) author's own intellectual creation. If there was some human intervention in the creation of the output, copyright protection is not excluded. In this respect, the creative process is decisive. A work that has been produced with the help of an AIS will only benefit from copyright protection if, during the creation process, the author has been able to express his creative abilities by making free and creative choices that stamp the work with his personal touch. If a work produced by an AIS contains 'sufficient traces of human creativity' in the creative process, it will thus be protected by copyright. Such output is then called 'AI-assisted output'. What degree of human creative intervention is required, is hard to determine. Creativity in machine-aided production may occur at three distinct (iterative) phases of the creative process, namely at the conception, execution and redaction phases.

In the conception phase, the human will often have the dominant role, as they will make most conceptual choices (the choice of the AIS, the selection and curation of input data, etc.). In the execution phase, the AIS takes over much of the human author's role. However, this does not mean that the user remains entirely passive during this phase. Often the user will monitor the output and give feedback to the AIS to guide it towards the desired output. Finally, in the redaction phase, the human author can make many additional creative choices, including rewriting, editing, formatting, cropping and refining. This is why mere human intervention in the conception and redaction phases is in many cases considered to be sufficient for copyright protection to arise ([1]). If the intervention of the user of a generative AIS is, however, limited to pushing a button for the AIS to operate, the output would arguably not be eligible for copyright protection, due to the absence of creative choices made by a human author. Much will depend of course on the facts and circumstances of each case.

It is thus clear that defining what the threshold is for human intervention to give rise to an original, copyright protected AI output is difficult. This should not come as a surprise, however, as AI output should be looked at on a spectrum that ranges from AI-assisted output to AI-generated output. At one end of the spectrum sit AISs that merely execute instructions given by a human author. Much like a regular photo camera, due to a lack of creative capabilities, these AISs cannot claim to be authors of the works they produce and any copyright on original works shall be owned by the author(s) of the works. Further along the spectrum, there are AISs that produce more creative outputs, which essentially still result from choices made by a human, by selecting the data input, modifying or selecting the output, etc. In such cases, copyright is still attributed to the human author who has been able to express their creative capabilities by making free choices. At the far end of the spectrum, one

([1]) See in this respect the CJEU's reasoning regarding machine-aided creation in the *Painer* case (C-145/10) and the decision of the French Court of Cassation regarding geographical maps directly created on the bases of unprotected satellite photographs in which the Court considered the maps to be copyright protected because they were 'the result of a personalised implementation of a complex technology by a process of transformation and improvement of choices, in particular colours, contrasts and of luminosity' (Cassation Civil I, 8 January 2002, *RIDA* 2002, No 193, 321).

could imagine futuristic, autonomous AISs that are capable of generating creative works that can no longer be distinguished from works produced by human authors. Currently, such AISs are considered not to exist (yet), but even if they did, as mentioned above, it would be debatable if they could ever be entitled to copyrights because they are not human and it could be argued that any ownership of copyright in their outputs would be reserved for the human that developed the autonomous generative AIS.

It is thus clear that only AI-assisted outputs (and not AI-generated outputs) are eligible for copyright protection. The question remains, however, as to who owns the rights in AI-generated and AI-assisted output. For AI-generated output, this question can easily be answered. As AI-generated output is not eligible for copyright protection, it becomes part of the public domain. This implies that the output can be freely used by anyone and no one owns the rights. For AI-assisted output, the question of ownership of copyrights is, however, far less clear. As mentioned above, copyright vests in the author or authors of the protected work, i.e. the person(s) whose intellectual creation the work(s) express(es). For AI-assisted outputs though, it is hard to establish whose intellectual creation the work(s) express(es). Are the programmers and designers of the generative AIS to be considered as authors? Or is the author the AI's trainer who has selected and curated the AIS's input? Could it perhaps be the user of the AIS who has selected and modified the output? Or should it be the investors of the AIS, the general public or even the government that owns the copyright in AI assisted output? As mentioned above, creativity in machine-aided production may occur at several stages of the creative process and the personality of multiple people may be reflected in the final work. As such, it could even be argued that copyright in AI-assisted output should be co-owned by several of the people mentioned above. Presently, however, no clear solution to this question exists. A one-size-fits-all solution would probably be hard to find, and a case-by-case assessment of the ownership question will be required instead. In the meantime, however, this issue will likely be solved through contractual agreements between the relevant parties. Indeed, in practice, the general terms and conditions of generative AI tools often include provisions on the ownership of outputs which clarify whether the AIS providers claim any rights in the output generated.

3. Artificial intelligence outputs and copyright infringement

Copyright cuts both ways, of course. Whereas certain AI outputs may be copyright protected, other AI outputs may constitute a copyright infringement. Indeed, much like any other content, AI outputs that resemble a pre-existing work may implicate copyright infringement. Considering that 'style' is not protected, generative AI outputs that resemble the style of a famous artist (e.g. Johannes Vermeer or Andy Warhol) do not necessarily amount to copyright infringement. However, as soon as a substantial part of a copyright protected work is copied in an AI output without permission of the right holder and without an exception to apply, copyright infringement will exist. Put simply, copyright infringement may exist if generative AISs produce outputs that copy one or more original elements of existing copyrighted works. This means that if a generative AIS inadvertently copies a part of an original work, this will be deemed an infringement of the reproduction right if the part that is reproduced is considered 'the author's own intellectual creation'. That is why some generative AI tools decline prompts that ask for output 'in the style of' a living artist. The reason for this is of course that, in such

cases, there is too high a risk that original elements of the pre-existing works of such artists would be copied in the output. Moreover, given that the resemblance with the pre-existing work would most likely stem from the fact that the AIS was trained on that work, it would be hard to argue that the output concerns a fully 'autonomous creation' that was created without knowledge of the pre-existing work and that, therefore, no copyright infringement should be withheld.

When AI outputs do copy elements which are the expression of the intellectual creation of the author of a pre-existing work, such AI outputs will most likely have to be considered an 'adaptation' of the pre-existing works, requiring the authorisation of the author(s). Indeed, as long as the form of the original work remains recognisable, any reproduction in which the original is substantially altered is subject to the exclusive reproduction right of the author(s). In addition, such AI outputs may also infringe the moral rights of the author of the pre-existing work, such as the right to paternity and the right to integrity.

If copyright infringement is confirmed, the question then arises as to who is liable for such infringement. When looking at infringements resulting from AI-assisted outputs, the answer is relatively straightforward. By definition, AI-assisted outputs imply some kind of human intervention. As a consequence, it may be assumed that the infringement is caused by the user either using the generative AIS for an unlawful reproduction or failing to sufficiently verify the AIS to avoid copyright infringement, so they will be liable. Indeed, a cautious user of a generative AIS would do well to verify if the use of a certain generative AI tool may lead to copyright infringements, for example, given the training data used. At very best, the user could try to shift the liability to AIS provider if it can be proven that the infringement can be blamed on this provider. For AI-generated outputs that constitute copyright infringement, allocating liability is harder. In the absence of legal personality, AISs that autonomously generate infringing outputs, cannot be held liable. As such, a fault-based liability regime is not meaningful for this situation and another culprit must be found. Who this should be in the case of autonomous systems has not yet been decided. The question of liability for damage caused by autonomous systems is however not specific to copyright. For example, self-driving cars may also cause damages, for which liability is difficult to allocate. As this topic is still evolving, we limit ourselves to reporting that the solutions that are currently on the table are a product liability regime, an objective liability regime or a *sui generis* regime where liability is placed on the party that can best mitigate the risk of infringement, combined with compulsory insurance.

4. The future of generative artificial intelligence and copyright

The proliferation of generative AI outputs raises many questions about what lies ahead. From a policy perspective, it could be questioned if in the future certain AI-generated outputs should be awarded copyright protection, and whether this would create benefits for the European Union's AI industry. Looking back at the rationale behind copyright protection, i.e. to encourage the creation of original works, copyright protection for AI-generated outputs seems counterintuitive. Indeed, it is hard to conceive why AISs (as machines) would need to be incentivised to produce more 'creative' outputs. Moreover, awarding copyright protection to AI-generated outputs could also bring about unwanted side effects, such as a sudden increase in AI-generated creations which could hypothetically even lead

to all reasonably possible creations being exhausted at some point (e.g. a sufficiently powerful AI could rapidly generate every melody and rhythm that is likely to be reasonably appealing to human ears in a matter of hours, thus 'exhausting' copyrights on music, since future works would then infringe the AI's copyrights to its creations). This would not only elevate the risk of copyright infringement, but could potentially also make it harder for humans to create original works. Essentially, awarding copyright protection to AI-generated outputs runs the risk of putting human creators in a position where they are no longer able to compete with AISs.

Alternatively, the possibility of establishing a related right or *sui generis* right for AI-generated outputs could be considered, with a view to protecting the investment made in AISs and incentivising further research into AISs. Related (or neighbouring) rights are rights related to copyright that provide persons or entities operating in creative industries that are not creators of copyright-protected works with a protection similar to copyright. The major difference between related rights and copyright is that related rights do not require originality or authorship. That is of course what makes related rights interesting for AI-generated output. Related rights already exist for databases in the EU, and for phonogram producers, broadcasters, film producers, publishers of press publications, etc. It could, for example, be argued that AI produced audio output could be considered a 'phonogram', so users of a generative AIS that triggered the act of fixation of sounds by activating the AIS are to be considered 'phonogram producers' that enjoy a right of reproduction, distribution and communication to the public of the AI produced output. Likewise, some consider establishing a new *sui generis* right for AI-generated outputs that covers all kinds of outputs (and not only those outputs covered by already existing related rights). However, such rights would presumably still lead to a strong increase in the amount of AI outputs being generated, which may potentially put pressure on the creation of original human works.

In this respect, it should be noted that, authors are already expressing concerns over a loss of income due to them being replaced by machines in sectors ranging from writing to music and visual arts. Indeed, generative AISs are able to churn out 'creative' works much faster and cheaper than human authors, thus demonstrating a great potential to strongly disrupt the market for human literary and artistic works. However generative AISs are only capable of doing so because they are essentially piggybacking on previous human creations that served as their training materials. This has triggered calls to compensate human authors for their loss of market share and income, for example by introducing an AI levy system that requires providers of generative AISs to pay a remuneration for producing output that has the potential to replace human creations, or by providing a system where human authors are compensated for the use of their works in AI training.

Presumably, this question of remuneration for human authors will have to be tackled together with the question of whether a separate IP regime for AI-generated outputs is needed in the future. Before introducing such new rights, someone has to carefully assess whether there is an actual market failure to be solved. Is there a risk that copying and freeriding on AI-generated outputs would result in less investment in AI outputs and less access to such outputs for the public, in a way that it would negatively affect the cultural, social and economic advancement of society? Currently, there appears to be no proof that a new IP regime would be required to incentivise the research and development in the field of AI. On the contrary, given that AISs often are protected by other intellectual property rights or other regimes, such as patent protection or trade secret protection, further legislative initiatives in this respect might not be necessary for the time being.

5. Conclusions

In conclusion, the evolving landscape of generative AI presents complex challenges at the intersection of creativity, ownership and copyright. As AISs continue to advance, their ability to produce outputs that mimic human creativity raises fundamental questions about the application of copyright law.

First of all, the question of the lawfulness of training generative AISs on copyright protected works has challenged the existing EU copyright framework and has even led to two new harmonised exceptions for text and data mining. Secondly, the outputs of generative AISs have reshaped the notion of creativity. Whereas it has been confirmed that generative AISs cannot be authors, it remains unsettled what degree of human creative intervention is required in order for outputs that are produced with the assistance of generative AISs to be eligible for copyright protection. And even if such copyright protection is awarded to certain outputs, the debate on who should own the rights in such outputs are still ongoing.

There is no controversy or uncertainty about the fact that generative AI outputs can infringe other copyright protected works. However, if such generative AI output were made without human intervention, it is not yet fully clear how liability for such copyright infringement would be dealt with. Lastly, looking at the future, important decisions will have to be made regarding the IP status of AI-generated outputs, and regarding the status and remuneration of human authors that may otherwise feel left behind. Balancing incentives for innovation with the rights and interests of human creators and society at large will thus be crucial in shaping the future of AI and copyright law.

A legislative attempt to reduce problems: the ambitions of the EU's Artificial Intelligence Act

1. Overview of the origins and principles of the Artificial intelligence Act

The Artificial Intelligence Act and its ambitions – context and background

Over the past few years, the importance of AI technologies and their potential impact has been increasingly recognised by policymakers all over the world. While AI offers many opportunities for society, it also presents risks if unchecked. For that reason, the EU has been devoting efforts to making sure that, within the EU, AI is used in a trustworthy manner and to the benefit of EU citizens.

In 2018, the *Coordinated Plan on Artificial Intelligence* was published. The plan was a joint commitment between the European Union, its MSs, Norway and Switzerland to maximise Europe's potential to compete globally, while making sure that AI technologies work for people and are a force of good in society, and that the EU is a place of excellence from lab to market. The plan was updated in April 2021 on the occasion of the European Commission presenting its AI package, which included the proposal for the AIA and the related impact assessment, in addition to a communication clarifying the intent of the EU to foster a European approach to AI

The AIA must hence be understood in this light of managing and mitigating any risks that AI may present to society, such as risks to the health, safety and fundamental rights (and freedoms) of natural persons, but also risks to public interests, which may include public health and economic safety interests, social interests, respect for private property and the protection of critical infrastructure. The ambition is to manage these risks appropriately, without creating an unnecessary burden that might unnecessarily stifle innovation.

In January 2024, the European Commission launched an AI innovation package to support AI startups and SMEs, as laid out in its communication on boosting startups and innovation in trustworthy AI.

At the time of writing, the AIA has been approved by the Council but not yet published. However, while the final text has not been published in the EU's Official Journal, a stable version was already approved by the European Parliament on 13 March 2024. Hence, the content of the future act was already largely clear, and this version has been used as the basis for this chapter. On 21 May 2024, the version that is expected to be published was shared in the press release announcing the approval by the Council. The changes made to the version of 13 March 2024 in the latest version, dated 14 May 2024, relate to numbering, structure, grammar and language consistency and do not influence the content of this analysis.

When will the Artificial Intelligence Act commence (material scope)?

In order to understand which instances of AI are regulated, it is important to have a look at the main definition of AISs to which the AIA applies. An AIS, as defined in Article 3(1) of the AIA, is 'a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments'.

This definition is very broad, meaning the AIA has a large reach. However, the AIA also outlines some situations that are exempt from its scope, such as:

- AISs and models, including their output, specifically developed and put into service for the sole purpose of scientific research and development;
- any research, testing and development activity regarding AISs or models prior to being placed on the market or put into service, unless the testing is done in real-world conditions;
- AISs released under free and open-source licences, unless they are placed on the market or put into service as HRAISs, prohibited AISs or AISs with specific transparency obligations (under Chapter IV of the AIA);
- natural persons using AISs in the course of a purely personal, non-professional activity.

General purpose AISs (GPAIS) and GPAI models (GPAIMs) are a specific subset of AISs, defined in the AIA. GPAIMs can serve a variety of purposes, when used directly or when integrated into other AISs, in particular because the models that they are based on display significant generality and are capable of competently performing a wide variety of tasks. GPT-4 is a commonly known example of a GPAIM. For such models, the AIA envisages some specific obligations because of their broad potential impact (see further under 'A risk-based approach to AI'). This is irrespective of any other rules that may apply, in particular when AISs are built on top of these GPAIMs.

The Artificial Intelligence Act – a risk-based approach to artificial intelligence

A distinct and important feature of the AIA is that it follows a risk-based approach. Risk under the AIA must be understood as 'the combination of the probability of an occurrence of harm and the severity of that harm'. The AIA aims to protect society against such harms, which might be material or immaterial, including physical, psychological, societal or economic. The values to be protected are the health, safety and fundamental rights (and freedoms) of natural persons, along with various public interests, which may include public health and economic safety interests, social interests, respect for private property and the protection of critical infrastructure.

The AIA deals with risk in two main ways:

- by determining risk categories that define whether an AIS is regulated or not;
- by imposing risk assessment, management and mitigation obligations on AISs that present a high level of risk.

This section covers the risk categories, other obligations related to risk are covered later.

The AIA divides AISs into four categories according to their potential risk level. These risk categories can be understood as a pre-determination by the EU legislator of the risk level that is inherently present in certain types of AI or certain uses of AI.

The **first category** is comprised of AI practices that pose an **unacceptable risk** and are therefore **prohibited**. This means that, for this category, the EU legislator has decided that the risk is inherently too high, even with potential safeguards to assess, manage and mitigate the risk.

Article 5 of the AIA provides a detailed list covering certain practices such as the use of AI for social scoring purposes, AI aimed at exploiting the vulnerabilities of persons due to their age, disability or a specific social or economic situation, or AISs that aim to create or expand facial recognition databases through the untargeted scraping of facial images from, for example, the internet or CCTV footage. However, to fall under this category, the practice must match the specific descriptions listed in Article 5. If it does not, the AIS may be permissible, albeit likely as an HRAIS.

The **second category** concerns AISs that pose a **high risk**. These are the AISs that are most directly targeted and regulated by the AIA. These systems are deemed permissible to use, but the AIA imposes strong risk assessment, management and mitigation measures on the use of such systems, to avoid such inherent risks from materialising. **HRAISs** cover the following.

- AISs that are safety components of products or are themselves a product that is already regulated by EU law listed in Annex II of the AIA and already requires a third-party conformity assessment, with a view to placing it on the market. For products covered by Section B of Annex II, the AIA does not apply directly, but the AIA's requirements for HRAISs will be indirectly added when the delegated acts are adopted in those sectoral laws.
- AISs listed in Annex III, which include:

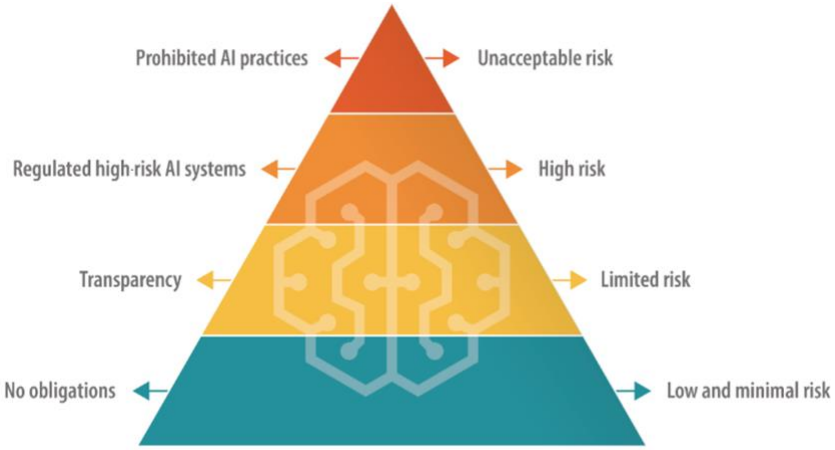
    a) remote biometric identification systems, emotion recognition systems, and biometric categorisation systems using sensitive or protected attributes;
    b) AIS systems used as safety components in critical infrastructure;
    c) certain AISs to be used in education and vocational training (e.g. access or admission to education, evaluation and assessment);
    d) certain AISs to be used in employment, the management of workers and access to self-employment, such as AISs for recruitment or performance evaluation;
    e) certain AISs to be used in relation to access to and the enjoyment of essential private services and essential public services and benefits;
    f) certain AISs to be used in law enforcement;
    g) certain AISs to be used in migration, asylum and border control management;
    h) certain AISs to be used in administration of justice and democratic processes.

For AISs in Annex III, an argument can exceptionally be made that they are not high risk, for example, when the AIS is meant to perform a narrow procedural task, or when the AIS is

merely intended to improve the result of a previously completed human activity. However, if the AIS involves the profiling of natural persons, this exception cannot be relied on.

The **third category** concerns AISs that pose **limited risk** to human safety and to fundamental rights, but where the main risks might arise from the fact that it may not always be clear that an AIS is involved. For this category, for example, relating to deep fakes or to AI that interacts with users directly, like chatbots, the AIA provides transparency requirements. This also covers many uses of GPAISs.

The **fourth** and last category concerns AI that poses **minimal risk or no risk at all**, which covers many AISs (e.g. AI video quality enhancer, AI spam filter). The AIA does not impose any requirements on these AISs (i.e. does not regulate them).



Source: European Parliament, briefing on the Artificial Intelligence Act

In addition to these main categories, the AIA also provides for specific rules relating to GPAIMs. These rules apply irrespective of the aforementioned risk categories because of the special nature of GPAIMs. Because of their widespread use and potentially broad impact, additional transparency measures relating to creating clarity about how the model works are in place for all GPAIMs.

All GPAIM providers will have to:

1. draw up and keep up-to-date technical documentation of the model, including its training and testing process and the results of its evaluation;
2. draw up, keep up-to-date and make available information and documentation to providers of AISs who intend to integrate the GPAIM into their AIS (i.e. for providers of AISs to be able to understand the capabilities and limitations of the GPAIM so that they can comply with the relevant obligations);
3. put a policy in place to respect EU copyright law;
4. draw up a sufficiently detailed and public summary of the content used for the training of the GPAIM.

In addition, providers of GPAIMs with systemic risk will also have to:

5. perform model evaluations;

6. assess and mitigate possible systemic risks at the EU level;
7. document and report information regarding serious incidents and possible corrective measures;
8. ensure an adequate level of cybersecurity.

It should be noted here already that, despite the detailed rules, many organisations that intend to use AI will not be regulated under the AIA. This is because all categories other than minimal risk are worded in a specific way. Many use cases will not fall under any such description and will therefore land in the residual category of minimal risk. However, this does not mean that such AI applications entail no risk, only that the AIA does not regulate them. Organisations should still assess the risks of an AI project in any case, as part of good practice, but usually also in the light of obligations imposed by other laws (e.g. data protection, sectoral rules), contractual or funder requirements, etc.

Regulated roles under the Artificial Intelligence Act

Not everyone that interacts with AI is subject to the AIA. The act's personal scope – i.e. the persons for whom the act creates rights or obligations – is limited to **operators**, a notion that aims to cover the relevant stakeholders in the value chain of AI products.

- **Providers** are the most regulated stakeholders, because of the decisive role they have in the design and development of an AIS. Providers are the natural or legal persons, public authority, agency or other body that develops or has developed an AIS and **places them on the market or puts them into service** under their own name or trademark, whether for payment or free of charge.

- **Deployers** are the natural or legal persons, public authority, agency or other body **using** an AIS under their authority, except when the AIS is used in the course of a purely personal, non-professional activity.

- **Importers** are the natural or legal persons that are established or located in the EU **that place on the market** AISs that bear the name or trademark of a natural or legal person established outside of the EU.

- **Distributors** are the natural or legal persons in the supply chain, other than the provider or the importer, that **make** an AIS **available** on the EU market.

Note that a key difference between a provider and a distributor resides in their interaction with the market. Indeed, **providers** 'place on the market' (and put into service, which is notably broader) whereas **distributors** 'make available on the market'. The former relates to the first time an AIS is made available on the market, whereas the latter refers to the supply of AI products and systems in the course of a commercial activity.

The following example may illustrate what a common course of action could look like according to the AIA.

1. A provider develops an AIS and puts its trademark on it.
2. The provider makes it available for distributors by 'placing [the AIS] on the market'.
3. The distributors will then sell the AIS to retailers or directly to customers as a commercial activity. This constitutes 'making [the AIS] available on the market'.
4. Those who purchase the product from the distributors or retailers and use it will be considered as deployers.

Note however, that the AIA does not require such a reality. Providers may also exist where there are no importers or distributors, and even when there are no deployers. In fact, a provider can become a provider by putting an AIS into use, i.e. for its own use, without it being placed on the market and may even offer that AIS as a service to users directly.

These roles are most relevant in relation to HRAISs, which is the most regulated category of the AIA. Provider and deployer obligations also apply to limited-risk AISs (transparency) and specific provider obligations apply to GPAISs. However, the focus in what follows is on HRAISs.

Notably, roles in relation to HRAISs may change in practice. The AIA envisages certain situations in which any distributor, importer, deployer or other third-party may become requalified as a provider of an HRAIS, and thus become subject to the more extensive provider obligations of the AIA, namely when they:

- put their **name or trademark** on an HRAIS that has already been placed on the market or put into service;
- make a **substantial modification** to an HRAIS that has already been placed on the market or put into service in a way that remains high risk;
- modify the **intended purpose** of an AIS, including a GPAIS, which has not been classified as high risk and has already been placed on the market or put into service in such a manner that the AIS becomes an HRAIS.

In other words, if any operator reappropriates an HRAIS that has already been placed on the market or put into service, they will become a provider of HRAISs.

The notion of a 'substantial modification' relates to a change within the HRAIS itself, after it has been placed on the market or put into service, that was not anticipated or planned in the initial conformity assessment made by the provider, in a way that it remains an HRAIS and that its compliance with the HRAIS regulations or its intended purpose is affected (e.g. an important modification of the algorithm).

A change of purpose refers to a reorientation of the intended purpose of the HRAIS, regardless of whether the HRAIS itself has been changed or not.

The Artificial Intelligence Act – territorial scope

With regard to territorial scope, the AIA covers the use of AISs in the EU, specifically with the aim of guaranteeing trustworthy AI for EU citizens and persons located in the EU. To this extent it covers:

- providers placing AISs on the market, putting AISs into service or placing GPAIMs on the market in the EU, irrespective of whether those providers are located within the EU or in a non-EU country;
- providers of AISs that have their place of establishment or are located in a non-EU country, where the output produced by the system is used in the EU;
- deployers of AISs that have their place of establishment or are located within the EU;
- deployers of AISs that have their place of establishment or are located in a non-EU country, where the output produced by the system is used in the EU;
- importers and distributors of AISs when they operate on the EU market;
- product manufacturers placing an AIS on the market or putting it into service together with their product and under their own name or trademark.

Provider obligations under the Artificial Intelligence Act

As noted, HRAISs are the most regulated category of the AIA. Therefore, the most important provider obligation in the AIA relate to HRAISs.

As such, HRAIS providers must comply with a general set of requirements for HRAISs which are meant at least in part to assess, manage and mitigate risk, i.e. they must provide for:

- the establishment of a risk management system;
- the implementation of appropriate data management and data governance practices;
- the creation and maintenance of technical documentation on the HRAISs;
- the incorporation of automatic record-keeping/logging capabilities within the HRAISs;
- HRAISs to be designed and developed in a way that is sufficiently transparent for deployers to be able to interpret them correctly, providing information to this extent (i.e. to enable understanding and provide clear instructions);
- the effective possibility for humans to oversee the HRAIS while it is being used;
- HRAISs to be designed and developed in such a way that they achieve an appropriate level of accuracy, robustness and cybersecurity, and perform consistently in those respects throughout their life cycle.

Whereas the abovementioned obligations generally concern the development and design of the HRAIS, there are a number of additional obligations aimed at providers complying with the regulation. These obligations include the following actions:

- identify the HRAIS, notably by putting their name or trademark on it;
- maintain proper documentation for at least 10 years after the HRAIS was placed on the market or put into service;
- put a quality management system in place, from which several additional requirements stem;
- keep the logs that are automatically generated by the HRAIS;
- register the HRAIS in the EU database;
- ensure that the HRAIS undergoes a conformity assessment, draw up an EU declaration of conformity and affix the CE marking to the AIS (i.e. the standardised *conformité européenne* marking, which has been used since 1993 to designate products and services that comply with European quality standards), to indicate compliance with the AIA;
- set up a post-market monitoring system in a manner that is proportionate to the nature of the AI technologies and the risks of the HRAIS and make sure a feedback loop is created, for example, to ensure that any new identified risks are included in the risk management;
- notify the competent authority if the HRAIS presents a risk to the safety, health or fundamental rights of humans, and take the necessary corrective actions;
- report serious incidents to the market surveillance authorities;
- cooperate with the competent authorities and, upon request, demonstrate the conformity of the HRAIS;
- for providers located outside of the EU, appoint an authorised EU representative.

One of the main outcomes of the provider obligations for HRAISs is that a conformity check is done, an EU declaration of conformity is drawn up and CE marking is affixed before the system is put on the market, so that deployers, users and natural persons can see that the provider warrants compliance of the AIS with the AIA. Strong sanctions help ensure that providers do not take this obligation lightly.

Whether the conformity assessment has to be carried out by a third party depends on the type of HRAIS. The procedure to be followed is defined by the AIA in Annex VI (based on internal control) and Annex VII (based on third-party control). For now, most HRAISs will be assessed internally by the provider, with the possibility for the Commission to adopt delegated acts in the future to extend the types of HRAIS that must be subjected to third-party assessments, taking into account that those third-party capacities must also be developed. The AIA also gives strong importance to the role of standards in the field, which may create presumptions of conformity with the AIA, along with common specifications (which are a measure to replace standards in areas where standardisation has not happened, despite the Commission's requests) and codes of practice (which are a tool to facilitate compliance with the AIA for GPAIs).

Other provider obligations that should be noted are those for AIS providers that directly interact with natural persons (Chapter IV of the AIA) and for providers of certain GPAIMs, where transparency obligations apply. Another set of obligations are those that apply to providers of GPAIMs with systemic risk. For such models, additional obligations apply, relating to evaluation of the model, risk assessment and mitigation, documentation and cybersecurity. The AIA provides requirements for what must be considered a model with systemic risk.

Note that a GPAIS built on top of a GPAIM, whether with systemic risk or not, can constitute an HRAIS (or a minimal risk system, for that matter). Hence, AISs built on top of GPAIMs may incur HRAIS obligations as well.

Deep dive into provider obligations for high-risk artificial intelligence systems: data management and data governance

An AIS needs data to function. Processing data, finding hidden patterns, creating new data and producing new outputs are all evident parts of an AIS. However, this also means that the input data must be appropriate for AI processing – even more so when HRAISs are involved – to avoid undesirable outcomes. Therefore, the AIA requires that **training, validation and testing data sets** for HRAISs must be subject to **appropriate data government and management practices according to the intended purpose of the AIS**. These practices must in particular deal with:

- the relevant design choices;
- the data collection processes and origin of data, and for personal data, the original purpose of the data collection;
- the relevant data preparation processing operations (e.g. annotation, labelling, cleaning);
- the formulation of assumptions;
- an assessment of the availability, quantity and suitability of the required datasets;
- an examination aiming to identify possible biases that are likely to negatively affect human health, safety and fundamental rights, or lead to discrimination;
- appropriate measures to identify and prevent such biases;
- the identification of relevant data gaps or shortcomings that prevent compliance with the AIA, and measures to address those issues.

Moreover, training, validation and testing datasets must reach certain **quality standards**. They must be relevant, sufficiently representative, free of error and complete in view of the intended purpose (to the best extent possible), and have appropriate statistical properties. The datasets must also take into account the features that are specific to the environment, i.e. the geographical, contextual, behavioural or functional setting within which the AIS is intended to be used.

In relation to **special categories of personal data under the GDPR** (e.g. medical data, political opinions) the AIA creates a specific legal basis for the processing of such data, to the extent that it is strictly necessary for the purposes of ensuring bias detection and correction in relation to HRAISs. This is important, as such a situation was not comprised amongst the exceptions provided by Article 9, paragraph 2, of the GDPR, so a specific legal ground in the AIA was needed. In this case, however, several conditions must be met in addition to the general provisions set out by the GDPR (e.g. the data processing principles):

- the bias detection and correction cannot effectively be fulfilled by processing other data, including anonymised or synthetic data;
- the processed data is subject to strict organisational and technical safety measures to ensure that they remain secure, to avoid misuse and to ensure that only authorised persons have access to this data with appropriate confidentiality obligations being in place;
- the processed data may not be transmitted to or accessed by other parties;
- the processed data must be deleted once the bias has been corrected or once the data has reached the end of its retention period;

- the processing of such data must be justified in the records of processing.

When HRAISs are being developed without the use of techniques involving the training of models with data, the aforementioned rules shall only apply to the testing datasets.

Deep dive into provider obligations for high-risk artificial intelligence systems: the risk management system

Another important obligation for HRAIS providers that deserves some further elaboration is the requirement to establish, implement, document and maintain a risk management system.

The risk management system is a continuous iterative process that must be planned and executed at regular intervals (i.e. with regular systematic review and updating) throughout the entire life cycle of an HRAIS. A risk management system has the following steps:

- the identification and analysis of the **known and the reasonably foreseeable risks** that the HRAIS can pose to people's health, safety or fundamental rights when it is used in accordance with its intended purpose;
- an estimation and evaluation of the risks that may emerge when the HRAIS is used in accordance with its intended purpose but under conditions **of reasonably foreseeable misuse**;
- an evaluation of other possible risks based on the analysis of data gathered from the post-market monitoring system;
- The **adoption of appropriate and targeted risk management measures designed to address the risks identified** (specifically the reasonably foreseeable risks). The result should be that the relevant residual risk associated with each hazard and the overall residual risk of the HRAIS are judged to be acceptable.

The focus of the risk management process is on the risks that can reasonably be mitigated or eliminated through the development or design of the HRAIS, or the provision of adequate technical information.

When deciding on appropriate risk management measures, the provider must eliminate or reduce identified risks as far as technically feasible through adequate design and development of the HRAIS; where appropriate, implement adequate mitigation and control measures for addressing risks that cannot be eliminated; provide information and, where appropriate, training to deployers. Due consideration must be given to the technical knowledge, experience, education and training to be expected on the side of the deployer and the presumable context in which the system is intended to be used.

HRAISs must be tested for the purposes of identifying the most appropriate and targeted risk management measures, taking into account that the HRAIS must perform consistently for their intended purpose and remain in compliance with the AIA.

Testing may be done at any point during the process, but at the latest before placing the HRAIS on the market. Testing must be performed against metrics that were defined beforehand and probabilistic thresholds that are appropriate to the intended purpose of HRAIS.

When the HRAIS is intended to be used with minors or vulnerable groups, due consideration must be given to this when implementing the risk management system.

Deployer obligations under the Artificial Intelligence Act

As for provider obligations, the most relevant obligations for deployers are found in relation to HRAISs.

HRAIS deployers must:

- take appropriate technical and organisational measures to ensure that they use such systems in accordance with the instructions of use accompanying the systems;
- assign human oversight to natural persons who have the necessary competence, training and authority, along with the necessary support; in particular when the deployer exercises control over the HRAIS;
- ensure that input data is relevant and sufficiently representative in view of the intended purpose of the HRAIS, to the extent that the deployer exercises control over the input data;
- monitor the operation of the HRAIS on the basis of the instructions of use and, when relevant, inform providers of issues, and inform others (depending on the case: the distributor, importer or market authority) in the event of serious risks or incidents;
- keep the logs automatically generated by the HRAIS to the extent that such logs are under their control for a period appropriate for the intended purpose of the HRAIS, of at least 6 months;
- carry out a DPIA, if required by applicable data protection law, using information provided to them by the provider under its transparency obligation;
- inform natural persons of the fact that they are subject to the use of the HRAIS in cases covered by Annex III, where the AIS makes decisions or assists in making decisions related to natural persons;
- cooperate with national competent authorities.

While clearly of a secondary nature to the obligations of the provider, the obligations of the HRAIS deployer still clearly require that an organisation organises itself to comply fully with the AIA.

Moreover, in certain cases deployers of an HRAIS will have to carry out a fundamental rights impact assessment before being able to use the HRAIS system. This is the case in relation to HRAISs mentioned in Annex III for:

- deployers that are governed by public law;
- deployers that are private operators providing public services;
- operators deploying HRAISs intended to evaluate the creditworthiness of natural persons or establish their credit score (with the exception of AISs used to detect financial fraud); or AISs intended to be used for risk assessment and pricing in life and health insurance.

However, the obligation does not apply to point 2 of Annex III (AISs as safety components for critical infrastructure). The AIA provides for the main elements of such an assessment, and makes an explicit statement that the AI Office (a body established within the Commission that will support the AIA's implementation from a practical point of view) shall develop a template questionnaire for this purpose. A notable element is that the AIA also mentions that, where a DPIA already covers elements of this assessment, both assessments shall be conducted in conjunction with each other.

Another limited set of deployer obligations relating to the transparency of systems directly interacting with natural persons can be found in Chapter IV of the AIA, which concerns AIS deployers using deep fakes, AISs that generate or manipulate text which is published with the purpose of informing the public on matters of public interest, or AISs involving an emotion recognition system or a biometric categorisation system.

Enforcement and fines

The AIA sets up a detailed system of post-market monitoring, information sharing and market surveillance, along with enforcement measures. Each MS will have to establish at least one national authority that will act as a market supervision authority and be responsible for the supervision and monitoring of AISs after they have been placed on the market or put into service, and at least one national authority that will act as a notifying authority, which will be competent to designate and oversee conformity assessment bodies (notification bodies) in the context of pre-market AIS obligations (which is relevant when a third party conformity assessment is required).

The authorities established by MSs will have a wide array of means to control and oversee the market. Notably, they will be able to lead investigations, receive full access to the documentation and information of an AIS, evaluate the compliance of AISs with the AIA, and order corrective measures or withdraw/recall AISs that do not comply with the AIA.

Moreover, the AIA outlines the fines that can be imposed, should the AIA be violated. These fines target three types of stakeholders and vary depending on the severity of the violation.

- With regard to **AIS operators**, MSs are tasked with laying down their own rules on penalties, while staying within the framework of the AIA. As such, AIS operators can be subject to: a fine of up to EUR 35 000 000 or 7 % of their total worldwide annual turnover for the preceding year (whichever is higher) for violating the rules which relate to prohibited AI practices; a fine of up to EUR 15 000 000 or 5 % of their total annual turnover for the preceding year (whichever is higher) for certain violations related to HRAISs, to certain requirements related to notification procedures, and to certain transparency obligations; a fine of up to EUR 7 500 000 or 1 % of their worldwide annual turnover for the preceding financial year (whichever is higher) if the operator were to supply incorrect, misleading or incomplete information to the competent bodies and authorities. For small and medium-sized enterprises and start-ups, the fine is determined based on the lowest amount of the two possible thresholds.
- **GPAIM providers** can be fined by the Commission for an amount that does not exceed 3 % of their worldwide turnover for the preceding financial year or EUR 15 000 000 (whichever is higher).
- Lastly, **EU institutions, bodies and agencies** may be fined by the European Data Protection Supervisor for an amount that does not exceed EUR 1 500 000 if the violation is related to prohibited AIS practices, or EUR 750 000 if any other obligation or requirement relating to AISs is violated.

When determining the amount of the fine, regulatory authorities will have to consider the nature, gravity and duration of the infringement, while taking into account the principles of proportionality and appropriateness.

The **right to lodge a complaint** is also worth noting, as it is unusually broad and grants standing to any natural or legal person that has 'grounds to consider that there has been an infringement' of the AIA.

2. What does the Artificial Intelligence Act mean in practice for open data ecosystems?

Understand your project and your role

Open data offers many opportunities for organisations to set up AI projects that build upon the available datasets, and/or that create new data, derive insights and set up new business strategies. However, depending on what type of AI is used and the specific goal and context of the project, different obligations will apply.

The definitions of 'provider' and 'deployer' are both quite broad. A provider does not necessarily need to put an AIS on the market; it is sufficient that they 'put them into service'. The AIA makes it clear that 'putting into service' means the supply of an AIS for **first use**, either directly to the deployer or for the provider's own use in the EU, for its intended purpose. An organisation may therefore become a provider even through its own internal use of an existing model that it has adapted to fit the organisation's specific needs or the specific needs of the project in which it intends to use open data to reach its goals.

A deployer is anyone who uses an AIS under their authority. Using an AIS under the organisation's authority is not very demanding; simply registering with a specific AI tool, if for professional use, is sufficient.

However, one must keep in mind that the AIA does not apply to:

- AISs released under free and open-source licences, unless they are placed on the market or put into service as HRAISs, prohibited AISs or AISs with specific transparency obligations under Chapter IV of the AIA (e.g. those having direct interactions with users: chatbots, deepfakes, emotion recognition, biometric categorisation, text manipulation when that text is mean to be published to inform the public on matters of public interest);
- AISs and models, including their output, specifically developed and put into service for the sole purpose of scientific research and development;
- any research, testing and development activity regarding AISs or models prior to being placed on the market or put into service, unless the testing is done in real-world conditions;
- natural persons using AISs in the course of a purely personal, non-professional activity.

Hence, an organisation using AI on open data sets purely for research purposes will not be covered by the AIA. Note that the AIA does not apply to AISs released under free and open-source licences, but it does apply to AISs released under paid open-source licences, and to specific AISs built on top of these by another organisation and put into use within that organisation. Hence, an organisation might take an existing open-source model (paid or not), modify or adapt it for their own use (without wanting to commercialise the model or AIS) and still be considered a provider under the AIA if they put it into use within the EU for purposes other than research or personal use (which are also not subject to the AIA).

In addition, even if the AIA applies as such, given the broad definition of AISs, many use cases for AI will not be covered by the substantive rules of the regulation. The AIA only forbids certain uses,

regulates high-risk uses and provides some specific transparency obligations for limited-risk uses that are user-facing, including for GPAI applications and AISs built on such GPAISs. It also provides some specific transparency obligations for the providers of such GPAIMs, to allow organisations interested in leveraging the power and ease of use of such applications to understand the models' capabilities and limitations.

Knowing and assessing your role under the AIA is therefore essential in order to determine if and how the AIA applies (or not) to your intended AI use case. Specific sectors must be more careful because they are more likely to veer into high-risk territory (e.g. education, employment, banks and insurance), whereas other sectors are less likely to incur any obligations, unless the AIS is client-facing.

When dealing with known HRAISs or in sectors that are implicated in the high-risk applications of Annex III, an organisation that is a deployer of such a system must take particular care to understand its own obligations and be mindful of the situations in which they themselves may become a provider of that system, such as when they modify the intended purpose of the system (including even GPAISs that were placed on the market as non-high-risk), make substantial modifications to the system (while it remains high risk) or put their name or trademark on the system (which may be done in a commercial context without realising the consequences).

Using open data in artificial intelligence applications

Generally, when using open data in AI applications, it is important to take into account the potential limitations of that data, according to the specific licence that applies, and any other legal requirements (including data protection and privacy expectations) that may apply. The use of AI may also present privacy risks that are not present when using this data without AI, especially when combining different sets of open data.

Generally, even when on organisation is not a provider of an HRAIS, and therefore not subject to the requirements explained above, it may still be useful, even for minimal-risk applications or applications that are not covered by the AIA (e.g. purely for research purposes), to consider the AIA's data management requirements anyway. These requirements apply to all types of data: training, validation and testing/input data.

It is important to think about and clarify within the organisation the following data governance and data management elements, taking into account the specific intended purpose of the AI application.

- Making the design choices of the AIS explicit.
- Defining the data collection processes and specifying the origin of data, and for personal data, the original purpose of the data collection.
- Identifying and preparing the relevant data preparation processing operations (e.g. annotation, labelling, cleaning).
- Formulating the assumptions for the intended AI application.
- Assessing the availability, quantity and suitability of the needed datasets, managing the licences applicable to different open data sets and ensuring licence compatibility.

- Examining possible biases that are likely to negatively affect human health, safety and fundamental rights, or lead to discrimination.
- Implementing appropriate measures to identify and prevent such biases.
- Defining processes to ensure that data are relevant, sufficiently representative, free of errors and complete in view of the intended purpose, have appropriate statistical properties and take into account the features that are specific to the environment, i.e. the geographical, contextual, behavioural or functional setting within which the AIS is intended to be used.

In order to implement this appropriately, organisations should put in place technical and organisational measures to make this operational.

This should realistically include a governance structure for the specific AI project, where responsibilities are clearly allocated. Disseminating a mission statement or clear message on the goals of the AI application throughout the organisation may also help easily communicate the main message to all persons involved. To minimise risks related to data, appropriate knowledge and training is of paramount importance.

Risk assessment and risk management of open data artificial intelligence use cases

Similarly to the preceding section, lessons may be learned from the risk management system that is an obligation for HRAIS providers in cases where the intended AI use case is not an HRAIS or is even minimal risk or not within scope of the AIA at all. Introducing AI in an organisation always entails a measure of risk that must be managed.

Therefore, a form of risk assessment and management will be useful for any organisation using AI, not only for providers and deployers of HRAISs, but also in instances where the AIA does not apply (e.g. for research) or where the AIA does not regulate (minimal-risk applications or limited-risk applications where there is no risk assessment mandated but only transparency). Even in such cases, risk is always present.

A form of risk assessment and management may therefore be required for the following.

- As part of general good practice to avoid potential liability and negative exposure, reputational damage, etc.
- As part of other processes, for example, due diligence, certification efforts, contractual requirements, when required by a funder.
- When mandated by data protection law: if the organisation is acting a controller, the GDPR or LED may likely require a DPIA. The AIA makes explicit reference to DPIAs for deployers that are mandated to conduct a FRIA.
- As part of compliance with national or sectoral rules (MS or EU).
- As part of a broader risk management process at the level of the organisation.

Risk assessment may in particular cover the following elements:

- a general risk assessment for the use of AI within the organisation;
- a risk assessment of the specific AI application;
- a FRIA of the specific AI application, if required by the AIA or other rules;
- a DPIA of the specific AI application, if required by data protection law.

Hence, there is no single concept of an AI impact assessment, but rather different elements of risk assessment that may come together in an AI application or project.

For all of the above elements, there are already sources that address them, and which may be used by organisations to prepare themselves. Examples include:

- for a risk assessment of AI, some standards are aimed specifically at managing AI risk, such as ISO/IEC 42001:2023, ISO/IEC 23894:2023 or the National Institute of Standards and Technology's freely available *Artificial Intelligence Risk Management Framework*; in addition,

general (cybersecurity) risk management standards may be relevant as well, along with other guidance, such as ENISA's guidance document on cybersecurity for AI;

- for a FRIA, guidance may be drawn from general human rights impact assessment methodologies and guidance documents;
- for DPIAs, guidance provided by national data protection authorities may be taken into account.

The importance of appropriate training and an AI governance framework within an organisation can hardly be overstated in this context. AI governance training and certification can already be found on the market.

Timeline of the Artificial Intelligence Act and expectations for the future

The AIA will enter into force 20 days after its publication in the Official Journal of the EU. The AIA will in principle become applicable 24 months after its entry into force. However, the AIA provides for many more detailed exceptions to this general rule. The following milestones may be particularly relevant for organisations:

- **6 months** after the entry into force – prohibited AI practices will become applicable;
- **12 months** after the entry into force – obligations for GPAIMs will take effect, except for GPAIMs that have been placed on the market before this date; MSs will have to appoint competent authorities; and there will be the first annual review by the Commission of the list of prohibited practices and of Annex III. The Commission has a delegated power to also add, remove or adapt use cases in the areas of Annex III (the areas will remain the same at that time, however);
- **18 months** after the entry into force – the Commission needs to have issued implementing acts creating a template for high-risk AI providers' post-market monitoring plan and have provided guidance on the practicalities of defining HRAISs by providing a comprehensive list of practical examples of high-risk and non-high-risk use cases for AISs;
- **24 months** after the entry into force – the AIA will enter into application, with most obligations taking effect from this date, including the obligations on HRAISs specifically listed in Annex III, which includes AISs in biometrics, critical infrastructure, education, employment, access to essential public services, law enforcement, immigration and the administration of justice;
- **36 months** after the entry into force – obligations for HRAISs listed in Annex II will take effect, as well as the obligations for GPAIMs that have been placed on the market before.

The AIA also has specific rules to deal with HRAISs that were on the market before the entry into force of the AIA. For most HRAISs, the AIA will start to apply as soon as those systems are subject to significant changes in their design. HRAISs intended for use by public authorities that were on the market before the entry into force of the AIA must become compliant within 6 years of the date of

entry into force (i.e. 4 years after the entry into application). For certain AISs that are components of the large-scale IT systems established by EU law in the areas of freedom, security and justice, such as the Schengen information system, 31 December 2030 is indicated as the date by which those systems have to become compliant with the AIA.

In addition to what was mentioned before, there are a number of topics that must still be addressed in the coming months and years:

- either by the Commission, to which the AIA delegates a number of tasks to be covered through:
    a) delegated acts (e.g. amendments to Annex III, amendments to the requirements for technical documentation of HRAISs),
    b) implementing acts (e.g. for approving codes of practice for GPAIs, operational rules for AI regulatory sandboxes, common specifications where standards do not apply), and
    c) guidance (e.g. on high-risk and non-high-risk use cases, the interpretation of the prohibitions, the application of the definition of an AIS, the concept of substantial modification to requalify operators in the AI value chain as providers);
- or by the European Artificial Intelligence Board, which can provide advice and issue recommendations and opinions;
- or by the AI Office, a body set up within the Commission, which will be tasked with various practical tasks to support the implementation of the AIA, including making standardised templates to support AIA compliance.

Hence, even before the application of the AIA 24 months after its entry into force, some guidance and clarifications are to be expected. The AI Office has already been established, which shows the clear intention of the EU legislator for this body to actively support the implementation (and preparation) of the AIA. This is not surprising, as the AI Office must support the drafting of codes of practice to aid in ensuring GPAIMs' compliance with the AIA, which the AIA requires to be ready no more than 9 months after the entry into force of the regulation.

3. Conclusion

With the AIA, the EU is taking a big step forward towards a future of trustworthy AI that will benefit society while making sure the EU's values, along with the health, safety and fundamental rights of its citizens and inhabitants, are protected.

At the time of writing, the AIA is yet to be published. Nonetheless, organisations wanting to leverage the power of AI in open data ecosystems should start preparing for its arrival, when the AI use case they have in mind is subject to regulation under the AIA. Even when this is not the case, the AIA provides valuable input on how to deal with data management and risk management, which must be addressed irrespective of whether the AIA applies to the use case or not. In that sense, staying up to

date with the evolutions surrounding the AIA is a good investment for all organisations aiming to leverage AI when using open data.

Currently, despite the AIA not yet being published, there is already an abundance of information available on the topic of AIA compliance, ethical compliance of AI, standards that might apply to AI risk assessments and to other assessments related to AI, such as FRIAs. It is not feasible for organisations, especially small and medium-sized enterprises working with open data, to process all this information.

Therefore, realistically, more official guidance and official templates are needed for organisations to be able to navigate this complex regulation. Such guidance is however to be expected from the Commission, the European Artificial Intelligence Board and the AI Office. In addition, it may be expected that the market will also add to creating services and support for operators to comply with the AIA. Certain standards, AI governance courses and certification already exist for early adopters. It can be expected that this trend will continue and that, with a combination of third-party services and official templates and guidance, AIA compliance will be increasingly facilitated in the coming years.

Overall conclusion on legal challenges in the intersection between artificial intelligence and open data

As this report shows, the legal framework for AI is still very much in flux, with several key questions still unresolved. It is clear that AIs can only be developed and used in compliance with the EU's fundamental rights framework, including in relation to data protection law; there is, however, a lack of good practices and real-life experience in how to deal with data protection issues and data subject rights requests. Similarly, it is fairly broadly accepted that AIs cannot be the beneficiaries of intellectual property rights, and that the creators of AIs should respect the copyrights to the source data that they use as training materials, but it is not yet fully established whether this implies that their consent should be obtained, or whether they are entitled to compensation. Lastly, it seems highly probably now that the EU will soon have its own legal framework specifically targeting AISs, using a risk-based approach that stratifies legal obligations depending on the risks of the AIS's use case, but it remains to be seen how this framework will affect EU and non-EU AI developers in practice, and how effective it will be.

In this rapidly evolving legislative environment, it is also difficult to predict how open data ecosystems will be affected. The potential opportunities are obvious: open datasets can serve as training materials for advanced AIs, thus leveraging an important European resource and contributing to the emerging AI economy. Furthermore, those advanced AIs can in turn be used to analyse datasets to identify new patterns and new knowledge, thus increasing the value of open data.

Nonetheless, the risks are equally clear. In the current state of play, it is not certain that advanced AI sets can be easily operated within the confines of European data protection law, or that intellectual property rights – including licence terms that can apply to open data – are appropriately taken into consideration. With those concerns in mind, the benefits of the interaction between AI providers on the one hand, and open data providers and the general public on the other hand, might be unbalanced in favour of the AI providers: they are able to access and build on large volumes of open data to create highly powerful and economically valuable AIs, whereas the benefits to data providers and the general public are currently still more nebulous.

The emerging AIA may prove to become a part of the answer to this conundrum, although it will also give rise to new challenges, including notably whether it will be effective as a risk management solution, and whether it will not place the European Union's AI providers at a comparative disadvantage compared to their peers outside of Europe.

In the meantime, the principal recommendation may be that open data providers should be mindful of the opportunities and risks that broader AI adoption will bring, and of the fact that openly available datasets may act as fuel for potent future AISs. This problem is not entirely new, since open availability of datasets has always implied a certain faith that users of the data would be willing to act lawfully, in accordance with published legal terms.

Meanwhile, however, it may be advisable for policymakers and data providers to reflect on the implementation of AI strategies, at the regional, national and EU levels, governing which datasets should be available as training materials, and under which conditions, and on how this should be monitored. Presently, in the fields of law and technology, few easy answers are available to open data providers. Yet, AI as a trend is here to stay, and AI-awareness is the first step towards managing the risks and harnessing the opportunities of an increasingly AI-driven society.

Bibliography

- Abbott, R. (ed.), *Research Handbook on Intellectual Property and Artificial Intelligence*, Edward Elgar Publishing Limited, Cheltenham, United Kingdom, 13 December 2022 (https://www.elgaronline.com/edcollbook/book/9781800881907/9781800881907.xml).
- Commission communication – artificial intelligence for Europe, SWD(2018) 137, COM(2018) 237 (https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN).
- Custers, B. and Fosch-Villaronga, E. (eds), 'Law and artificial intelligence – Regulating AI and applying AI in legal practice', *Information Technology and Law*, TMC Asser Press, The Hague, Netherlands, 5 July 2022 (https://link.springer.com/book/10.1007/978-94-6265-523-2).
- Hallinan, D., Leenes, R. and De Hert, P. (eds), 'Data protection and privacy, volume 13 – Data protection and artificial intelligence', *Computers, Privacy and Data Protection*, Bloomsbury Publishing Inc., New York, United States, 25 March 2021 (https://www.bloomsbury.com/us/data-protection-and-privacy-volume-13-9781509941759/).
- Janssens, M.-C. and Gotzen, F., 'Kunstmatige kunst – Bedenkingen bij de toepassing van het auteursrecht op artificiële intelligentie', *Auteurs en Media*, No 3, Larcier, Brussels, Belgium, 2020, pp. 323-342 (https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias2933893&context=SearchWebhook&vid=32KUL_KUL:Lirias&lang=en&search_scope=lirias_profile&adaptor=SearchWebhook&tab=LIRIAS&query=any,contains,LIRIAS2933893&offset=0).
- Nikolinakos, N. Th., 'EU policy and legal framework for artificial intelligence, robotics and related technologies – The AI Act', *Law, Governance and Technology*, Springer, Cham, Switzerland, 7 July 2023 (https://link.springer.com/book/10.1007/978-3-031-27953-9).
- Vanherpe, J., 'Paris Congress ALAI June 22-23, 2023 – Artificial intelligence, copyright and related rights – Answers Belgium', *Auteurs en Media*, No 1, Larcier, Brussels, Belgium, 2023, pp. 143–158 (https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias4133231&context=SearchWebhook&vid=32KUL_KUL:Lirias&search_scope=lirias_profile&adaptor=SearchWebhook&tab=LIRIAS&query=any,contains,LIRIAS4133231&offset=0&lang=en).
- Wischmeyer, T. and Rademacher, T., *Regulating Artificial Intelligence*, Springer, Cham, Switzerland, 10 December 2019 (https://link.springer.com/book/10.1007/978-3-030-32361-5).

Publications Office
of the European Union