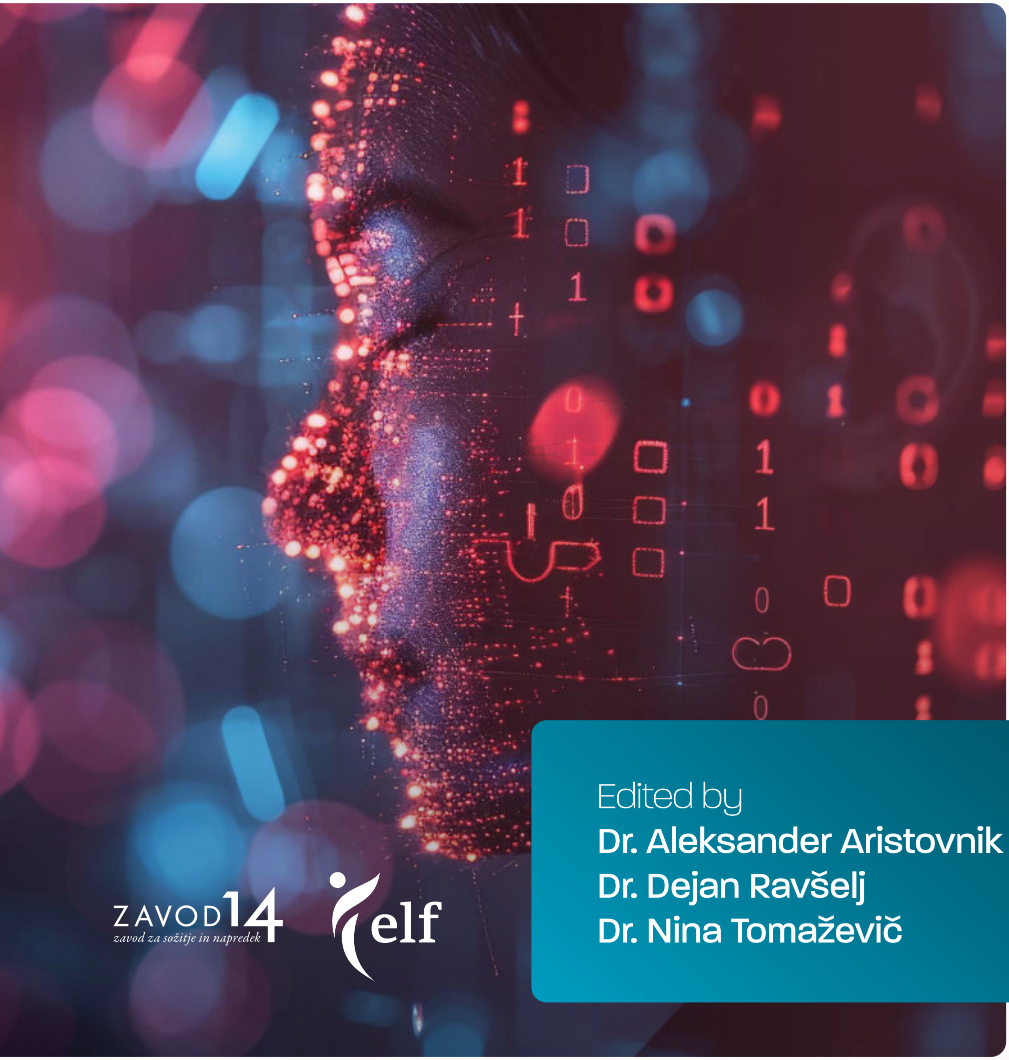
The Regulation of Artificial Intelligence: **Shaping Governance and Power Relations for a Liberal Future**



Edited by
Dr. Aleksander Aristovnik
Dr. Dejan Ravšelj
Dr. Nina Tomaževič

ZAVOD 14
zavod za sožitje in napredek

elf

THE REGULATION OF ARTIFICIAL INTELLIGENCE: SHAPING GOVERNANCE AND POWER RELATIONS FOR A LIBERAL FUTURE

Edited by Aleksander Aristovnik, Dejan Ravšelj and Nina Tomaževič

European Liberal Forum EUPF,
Rue d'Idalie 11-13, boite 6, 1050 Brussels, Belgium
info@liberalforum.eu
www.liberalforum.eu

Zavod 14,
Ljubljanska cesta 16, 3000 Celje, Slovenija
info@zavod14.si
www.zavod14.si

Graphic Design by spelabecaj.si,
freepik.com

ISBN: 978-2-39067-074-2 9782390670742

Published by the European Liberal Forum EUPF with the support of the Friedrich Naumann Foundation and Zavod 14. Co-funded by the European Parliament and ARIS (P5-0093). Neither the European Parliament nor the European Liberal Forum EUPF are responsible for the content of this publication, or for any use that may be made of it. The views expressed herein are those of the author(s) alone. These views do not necessarily reflect those of the European Parliament and / or the European Liberal Forum EUPF

European liberal forum (ELF)

The European Liberal Forum (ELF) is the official political foundation of the European Liberal Party, the ALDE Party. Together with 57 member organisations, across Europe we work to bring new ideas into the political debate, provide a platform for discussion, and empower citizens to make their voices heard. The ELF was founded in 2007 to strengthen the liberal and democrat movement in Europe. Our work is guided by liberal ideals and a belief in the principle of freedom. We stand for a future-oriented Europe that offers opportunities for every citizen. The ELF is engaged on all political levels, from local to European. We bring together a diverse network of national foundations, think tanks and other experts. At the same time, we are close to, yet independent from, the ALDE Party and other Liberal actors in Europe. In this role, our forum serves as a space for the open and informed exchange of views among a wide range of actors.

Zavod 14, zavod za sožitje in napredek

Zavod 14, zavod za sožitje in napredek is a non-profit (ELF full member) organisation that has its headquarters in Celje (Slovenia). Zavod 14 promotes social liberal ideas (balancing between individual liberty and social justice) and protects liberal values (e.g., democracy, the rule of law, social development, good governance etc.) The mission of Zavod 14 is to support civil society, integrate and cooperate with the interested public, transmit the perspectives of interested stakeholders to state and other institutions, cooperation in the preparation and implementation of politics, and contribution in joining the civil society initiative into international integrations.

Contents

Chapter 1 10

Legal Definitions of AI Systems: Insights from the AI Act
Đorđe Krivokapić, Andrea Nikolić, Ivona Živković

Chapter 2 32

Beyond the Hype: Navigating AI Regulation Through the GDPR and the AI Act in France and Europe
Sébastien Fassiaux

Chapter 3 54

The Italian Draft Law on Artificial Intelligence Systems: Key Differences and Complementarities with the EU AI Act
Gian Marco Bovenzi

Chapter 4 74

Regulation for Systemic Artificial Intelligence Safety: Interoperable Regulations and a Pragmatic Enforcement Strategy for Mitigating Unintended Consequences
Farhana Liza

Chapter 5 100

The EU Regulation of Generative AI: What Can We Learn from the Controversy Between the Italian Data Protection Authority and OpenAI
Michele Farisco

Chapter 6 120

Power Imbalances and the AI Act's National Supervisory Authorities: Bulgaria as a Case Study of the Challenges for the EU's Periphery
Fabio Ashtar Telarico

Contents

Chapter 7 142

The Regulatory Dimension of Artificial Intelligence in Romania
Melania-Gabriela Ciot

Chapter 8 170

We Do Not Want Big Brother in the EU
Šárka Shoup

Chapter 9 192

The Regulation of AI in Research by Adapting EU AI Regulations to African Contexts and Exploring Opportunities and Challenges for Harmonious AI Governance
David Mhlanga

Chapter 10 216

The AI Act Balancing Innovation and Regulation in Welfare: A Case Study of the e-Welfare System in Slovenia
Dejan Ravšelj, Matej Babšek

Chapter 11 244

Globally Applicable Gold Standard: Exploring the Extraterritorial Implications of the AI Act on International Arbitration
Jasmina Ibrahimpašić

Introductory remarks

Artificial intelligence (AI) technologies, including machine learning, deep learning, and natural language processing, are inherently global, cutting across national borders and necessitating international cooperation on AI legislation. Recognizing this, the European Union (EU) has taken a proactive stance in shaping a comprehensive AI policy framework designed not only to foster innovation but also to drive sustainable growth within its markets. By studying and comparing various policy approaches, the EU has identified key strategies that support responsible AI development, promote investment, and attract top-tier talent to its tech ecosystem. These efforts create a robust, competitive environment for AI researchers, start-ups, and established companies alike. Central to these initiatives is the EU's groundbreaking AI Act, which balances innovation with regulation by imposing stringent rules on high-risk AI applications while allowing greater flexibility in low-risk AI projects. This dual focus on accountability and transparency not only strengthens trust in AI systems across the EU but also sets the stage for a harmonized global approach to AI governance.

The primary aim of this publication is to offer an in-depth analysis of the wide-ranging implications of AI regulation within the European context. By critically assessing the current regulatory landscape, it sheds light on the impact of existing frameworks on AI development across the EU and its member states. The publication thoroughly examines discrepancies in national regulations, identifying areas where fragmentation could hinder harmonization efforts. These differences are juxtaposed with the recently adopted AI Act, offering a roadmap for achieving regulatory alignment. In doing so, the publication seeks to support the EU's broader objective of creating a unified regulatory environment, promoting both innovation and safety in AI technologies. Accordingly, the publication addresses the topic from three different perspectives.

The first perspective is focused on **legal definitions and foundations**. The **first chapter** provides an analysis of the definitions of AI systems, focusing on their core elements: inference, autonomy, and adaptability. It examines their significance in the context of the AI Act while distinguishing AI systems from other types of automated or smart systems. The **second chapter** examines the intersection of AI regulation with data protection law, particularly the General Data Protection Regulation (GDPR), and how AI systems have been subject to legal constraints for years. It discusses the role of France's data protection authority, the Commission Nationale de l'Informatique et des Libertés (CNIL), in enforcing regulations, the interactions between the AI Act and the GDPR, and the challenges of aligning competition, copyright, and privacy laws in regulating AI systems across various sectors. The **third chapter** examines the Italian draft law on artificial intelligence systems, highlighting its alignment with and harmonization with the EU AI Act. It addresses key areas such as health, labour, public administration, judiciary, cybersecurity, criminal law, and intellectual property, emphasizing the importance of responsible AI use, data protection, and promoting innovation while ensuring compliance with both national and European legal frameworks.

The second perspective is related to **European regulation and national perspectives**. The **fourth chapter** analyzes the evolving landscape of AI regulations, focusing on the challenges of ensuring systemic safety and pragmatic enforcement strategies while addressing unintended consequences such as stifling innovation, creating compliance burdens, and fostering international cooperation for consistent AI governance across different jurisdictions. The **fifth chapter** explores the controversy between the Italian Data Protection Authority and OpenAI regarding ChatGPT, highlighting the regulatory challenges and privacy concerns posed by generative AI systems. It focuses on data protection, transparency, risks of inaccurate information, and ethical issues related to user consent, ultimately discussing how these concerns intersect with broader European AI regulations like the AI Act and the GDPR. The **sixth chapter** examines the power imbalances created by the AI Act's framework for

national market supervisory authorities (MSAs), using Bulgaria as a case study to highlight the challenges faced by smaller and less affluent EU member states in effectively overseeing AI systems. It recommends policy changes to strengthen MSAs' capabilities and protect against regulatory capture by large technology companies. The **seventh chapter** discusses the regulatory framework for AI in Romania, focusing on national efforts to align with the EU's AI Act. It emphasizes the development of Romania's AI strategy, the integration of AI technologies across various sectors like education, healthcare, and agriculture, and the creation of a robust legal and ethical foundation to support innovation while protecting societal interests. The **eighth chapter** examines the EU's AI Act, focusing on the balance it seeks to strike between fostering innovation and ensuring the protection of fundamental rights. It explores the challenges and criticisms surrounding its implementation, particularly in relation to mass surveillance, biometric systems, and the potential economic impact on businesses, while highlighting the EU's effort to set a global regulatory standard for artificial intelligence.

The third perspective is focused on **global case studies and implications**. The **ninth chapter** explores how the EU's AI regulations can be adapted to the unique socio-economic, cultural, and technological contexts of African nations, addressing the challenges of applying a European framework to diverse African landscapes while proposing modifications to ensure that AI governance promotes innovation, ethical standards, and equitable benefits across the continent. The **tenth chapter** discusses how the AI Act is applied in the welfare sector, using Slovenia's e-welfare system as a case study to highlight how AI technologies are integrated into social welfare services. It addresses the challenges of balancing innovation with ethical concerns and emphasizes the need for transparency, accountability, and safeguards to protect vulnerable populations from potential AI misuse. The **eleventh chapter** explores the extraterritorial implications of the AI Act on international arbitration, examining how the regulation's broad scope may influence procedural rules, compliance obligations, and enforcement of arbitral awards. It draws parallels with the impact of the GDPR and assesses potential challenges for the international arbitration community in adapting to the new legal framework.

In terms of outcomes, the publication delivers crucial insights into the evolving landscape of AI regulation, highlighting its practical effects and the complexities of governing AI through multiple levels of oversight. It delves into the challenges associated with multilevel governance, where national laws often intersect or conflict with overarching EU policies. These insights provide a foundation for developing evidence-based policy approaches that reflect the diverse needs of EU member states while aligning with the strategic goals of liberal partners both within and beyond Europe. By addressing these regulatory challenges, the publication serves as a vital resource for policymakers, researchers, and industry leaders seeking to navigate the intricate dynamics of AI regulation, ensuring that innovation is encouraged without compromising ethical standards or public safety.

> ”
> The publication delivers crucial insights into the evolving landscape of AI regulation, highlighting its practical effects and the complexities of governing AI through multiple levels of oversight.

Chapter 1

Legal Definitions of AI Systems: Insights from the AI Act

Đorđe Krivokapić

Andrea Nikolić

Ivona Živković

1 Introduction

The emergence of new technology often raises the question of its regulation; namely, whether it is necessary to regulate such technology. There is a concern that technology is advancing more rapidly than legislators can effectively regulate it, leaving them struggling to predict all of its potential societal impacts. These dilemmas have also been present in the regulation of artificial intelligence (AI). Since arguments exist both in favour of and against regulation, the initial approach was to control AI through soft law mechanisms, including non-binding standards, ethical principles, and other guidelines. However, on the European level, it was decided to regulate AI by introducing a legislative instrument: the AI Act. Adoption of the final draft of the AI Act followed several months of intense negotiations due to conflicts over certain issues in the trialogue itself. Among others, controversial issues revolved around the definition of artificial intelligence itself, which is the focus of this paper.

We first analyse the definitions of AI systems proposed by various actors and stakeholders, examine their elements, and explain their significance before finally analysing the definition of AI systems used in the AI Act. While exploring the mentioned definition, we concentrate on its core elements: inferring, autonomy and adaptability. Inferring refers to the capability of AI systems to reason based on the input data they are given while autonomy refers to the ability of AI systems to operate independently, reasoning without human intervention. On the other hand, adaptability underscores AI's capacity to learn from new data and experiences, adjusting its behaviour so as to better fulfil its objectives over time. These capacities are crucial because they enable AI to navigate complex tasks and ultimately enhance efficiency and effectiveness across various industries, capabilities once traditionally exclusive to humans. Finally, we provide guidelines for distinguishing AI systems from information systems, computer software, automated decision-making systems, and smart systems, which are not within the scope of the definition.

2 Evolution of "AI System" Definitions in Negotiations of Adopting the AI Legal Framework (AI Act)

Defining AI is essential given that its definition determines the material scope of regulatory instruments and provides legal certainty. The absence of a clear legal definition of AI brings considerable challenges, including the inconsistent application of legal regulations, accountability issues, creating barriers to developing innovation, leading to economic uncertainties. Clear definitions offer understanding and predictability, harmonise the implementation of legal standards across jurisdictions, eventually strengthening the public's view of AI as trustworthy. This means establishing a universally accepted legal definition of AI is vital for ensuring accountability, fostering innovation and safeguarding ethical and legal standards concerning the development and use of AI.

Formulating a precise legal definition of AI proved to be a difficult task, bearing in mind the absence of consensus on its technical definition. First, challenges arise when defining intelligence itself, while the description that it represents the ability to behave or think intelligently is a vague and broad concept which can be analysed philosophically from different angles. One of these is that intelligence is characterised by rational behaviour: using all available information to evaluate and select the option that provides the greatest overall benefit (Marwala, 2017). An alternative approach could be to define AI by specifying the techniques and approaches considered as AI, although there is no universally accepted classification for AI techniques and approaches. Commonly listed specific techniques include machine learning, rule-based modelling, logic-based approaches, search and optimisation techniques, and genetic algorithms (Sarker, 2022), whereas some others classify AI as any technique used to achieve a specific objective, like computer vision, natural language processing, robotics, learning or reasoning (Norvig, Russell, 2020). Those classifications were used in negotiations on several international legal frameworks. However, various dilemmas and conflicts on adopting the ethical and legal framework were encountered during negotiations among European authorities, international organisations, and other stakeholders. In this chapter, we present the legislative history and explain the reasoning behind the proposed definitions.

Although the OECD AI Principles are a soft law instrument, they are the first intergovernmental standard on AI (adopted in 2019). These Principles are thereby the first relevant policy document to define AI systems, following consensus within a wide international forum. The very first definition contained in the Principles has already crystallised certain key concepts within the definition, which were later used by other policymakers, including the need to set relevant objectives (by humans), the reference to specific outputs, AI's influence on the environment, and the requirement for AI to be autonomous. During the negotiation process within the EU concerning the AI Act and other policy documents, reaching consensus on defining an AI system was a real challenge. The initial definition of the OECD as well as the revised definition published in November 2023, based on technological and market developments, provided necessary guidance and enabled informed debate, eventually impacting the definition incorporated into the EU's AI Act, along with a definition of AI included within the USA's Executive Order on AI.

Table 1: Comparison of the OECD's Definitions of AI Systems in 2019 (initial) and 2023 (revised)

Old OECD Definition of AI System (2019)	**New OECD Definition of AI System (2023)**
A machine-based system that can, for a given set of *human-defined objectives*, make predictions, recommendations, or decisions *influencing real or virtual environments.* AI systems are designed to operate with varying levels of *autonomy.*	A machine-based system 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.* Different AI systems vary in their levels of *autonomy and adaptiveness after deployment.*

Source: The OECD AI Principles.

Changes in the OECD's revised definition include doing away with the requirement that objectives must be defined by humans since an AI system itself can independently learn and set new objectives. Thus, the wording that AI systems can "infer how to generate outputs" based on the inputs they obtain not only from humans but from the environment was adopted. Besides the previously listed outputs that included predictions, recommendations or decisions, the updated definition states that AI systems can produce content as an output – referring to generative AI systems' ability to produce text, videos, audio, images and other kinds of content. Finally, a new element of the definition is adaptability. Namely, some AI systems, especially those utilising machine-learning techniques, can learn and adapt to new circumstances evolving beyond the design and deployment stages.

In June 2018, on the level of the European Union, the High-Level Expert Group on Artificial Intelligence (hereafter: AI Expert Group) was established by the European Commission. It includes 52 representatives from the public sector, industry and academia, together with independent experts. The expert group's main task was to analyse the need for the legal regulation of AI and to create an ethical framework applicable to artificial intelligence systems. The AI Expert Group published a few important policy documents. For example, in April 2019 it adopted the Ethics Guidelines for Trustworthy Artificial Intelligence along with other documents like the Assessment List for Trustworthy AI, the Policy and Investment Recommendations for Trustworthy AI, and the Sectoral Considerations on the Policy and Investment Recommendations, which together form the ethical framework for the application of the AI in the European Union.

The European Commission's proposed definition of AI emerged in 2018. It contained a reference to "intelligent behaviour", which was later omitted in legal definitions and AI's capacity to have some degree of autonomy in order to accomplish specific goals. The AI Expert Group followed that definition, scrutinised and upgraded it with explicit mention of knowledge-based and data-driven AI systems. Yet, at this stage the definition of AI & AI systems, although pivotal in outlining the capabilities of AI, were too descriptive and not easily comprehended in a legal context. This led to the definitions becoming significantly altered in the subsequent period.

In April 2021, the European Commission published its proposed AI Act. Afterwards, the lengthy trialogue between the European Commission, the European Parliament, and the Council of the European Union, considering guidance provided by the revised OECD definition, resulted in an agreement in December 2023. The table below shows the evolution of the definition of AI during the negotiation process as proposed by various bodies.

Table 2: Comparison of proposed definitions of "AI systems" in different legal regulations

European Commission's Initial Definition of AI (2018)	High-Level Expert Group on AI's Definition of AI (2018)
AI refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals. AI-based systems can be purely software-based, acting in the virtual world (e.g. voice assistants, image analysis software, search engines, speech and face recognition systems) or AI can be embedded in hardware devices (e.g. advanced robots, autonomous cars, drones or Internet of Things applications).	AI refers to systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to pre-defined parameters) to achieve the given goal. AI systems can also be designed to learn to adapt their behaviour by analysing how the environment is affected by their previous actions. As a scientific discipline, AI includes several approaches and techniques, such as machine learning (of which deep learning and reinforcement learning are specific examples), machine reasoning (which includes planning, scheduling, knowledge representation and reasoning, search, and optimization), and robotics (which includes control, perception, sensors and actuators, as well as the integration of all other techniques into cyber-physical systems).

AI Act's Initial Definition of AI System (2020)	AI Act's Final Definition of AI System (2024)
AI system means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. ANNEX I ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES referred to in Article 3, point 1: **Machine learning approaches**, including supervised, unsupervised and reinforcement learning, using a wide variety of methods including deep learning; **Logic- and knowledge-based approaches**, including knowledge representation, inductive (logic) programming, knowledge bases, inference and deductive engines, (symbolic) reasoning and expert systems; **Statistical approaches**, Bayesian estimation, search and optimization methods.	AI system means 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. ANNEX I ARTIFICIAL INTELLIGENCE TECHNIQUES AND APPROACHES referred to in Article 3, point 1 was not present in the final version.

Source: Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions on Artificial Intelligence for Europe, Brussels, 25.4.2018 COM(2018) 237 final; The European Commission's HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE, A DEFINITION OF AI: MAIN CAPABILITIES AND SCIENTIFIC DISCIPLINES, Definition developed for the purpose of the deliverables of the High-Level Expert Group on AI, Brussels, 18 December 2018; REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

The AI Act's initial definition of AI from 2020 shared some elements with the OECD's 2019 definition, including the human-defined objective, to generate outputs influencing the environment. However, the AI Act's initial proposal did not mention content as a possible output since concerns with the relevance and impact of generative AI systems emerged following the release of Chat GPT at the end 2022. Further, the AI Act's initial proposal did not refer to autonomy or adaptability, which were included in the final wording of the AI Act.

Further, Recital 12 of the AI Act reveals that the legislators' aim was to focus on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches that should not cover systems based on rules defined solely by people to automatically execute operations. A vital characteristic of AI systems is their capability to infer, which was incorporated in the definition. AI systems can infer how to generate outputs from the input and data they receive, yet also through learning from previously collected data using machine learning or from encoded knowledge or a symbolic representation of the task to be solved using logic- and knowledge-based approaches.

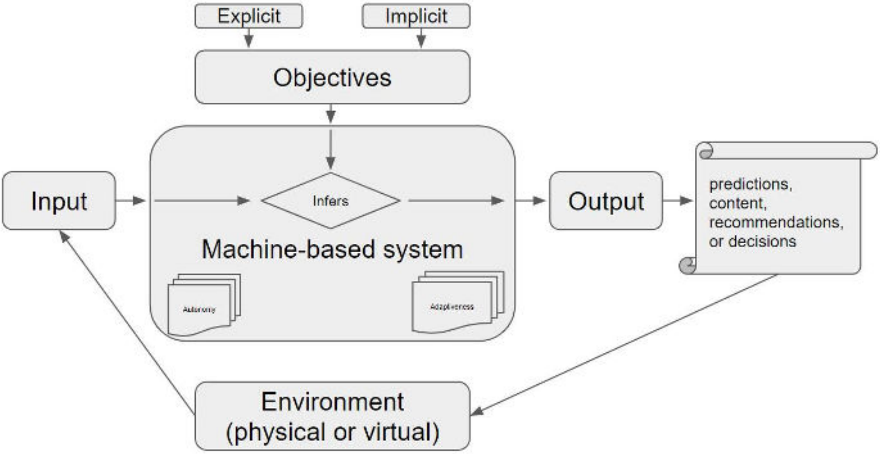
During the negotiations, the ultimate goal was to achieve legal certainty through a clear and concrete definition. The definition had to be neutral and flexible enough to cover all potential technological developments in the field. Thus, the final definition does not refer solely to systems designed by humans, nor software developed with one or more of the techniques including machine learning, logic- and knowledge-based approaches and statistical approaches listed in Annex I of the AI Act headed AI techniques and approaches. Annex I was not present in the final version of the AI Act, while the AI system is described as a machine-based system that can encompass all possible techniques and approaches. A machine-based system refers to the fact that AI systems run on machines. Namely, the final definition of AI system includes, albeit does not mention explicitly, different techniques like machine learning, knowledge-based approaches, and application areas such as computer vision, natural language processing, speech recognition, intelligent decision support systems, intelligent robotic systems and specific applications of these tools in different domains, or any other possible technique. Still, it is important to determine which technique would be used since the techniques AI systems use can influence the output but also produce different levels of risk. For example, the use of machine learning can create biases in data or algorithms, a lack of transparency in decision-making, leading to difficulties in accurately interpreting outcomes.

In the adopted version of the definition of an AI system, autonomy and adaptiveness became the main differentiating elements. The capacity of AI systems to operate with a certain level of autonomy namely enables them to act independently and function without human intervention. On the other hand, their adaptiveness, or self-learning capabilities, which are not mandatory elements of an AI system according to the AI Act definition, allows them to evolve and improve after the deployment phase, while in use. These crucial elements and their applicability across different fields are to be explored in more detail in the next section.

3 The Visual Model of AI Systems under the AI Act

Understanding the intricate structures and definitions of AI Systems as outlined in the AI Act is challenging, especially for interdisciplinary audiences, from technical to social sciences and humanities, who might be interested in its interpretation. Developing visual models may serve as a practical tool for demystifying these complexities. Their task is to provide a clear, graphical representation of how AI systems operate, including an illustration of the inputs, processes and outputs. By visually deconstructing the components and interactions of the AI system, these models enhance comprehension and facilitate effective communication across diverse fields of technology, law and policy. In this section, we develop these visual models in order to foster a deeper, more accessible understanding of AI systems and their core elements under the AI Act.

Diagram I: Basic Visual Model of AI Systems under the AI Act



Source: Authors' own work and design following the AI Act and the OECD's definition of AI system

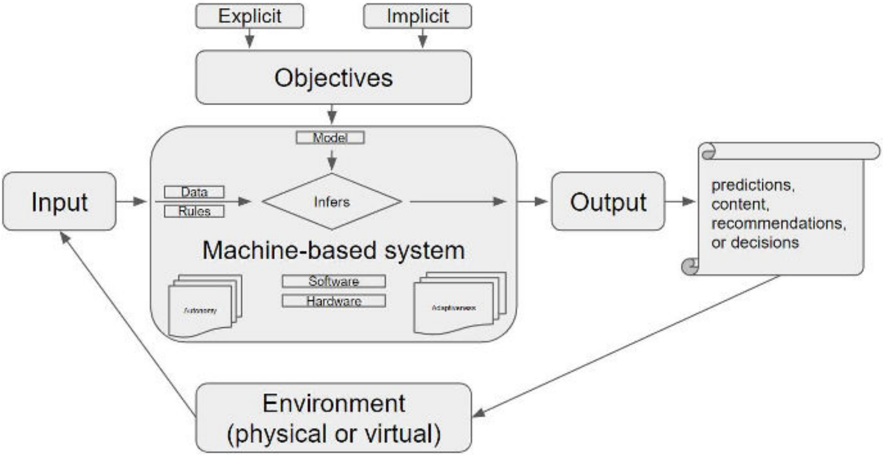
Diagram I provides a conceptual overview of an AI system as defined by the AI Act, illustrating the relationships between the various components and their functions.

1. **Objectives:** An AI system operates based on both explicit and implicit objectives. Explicit objectives are predefined by humans, while implicit objectives can be learned and set by the system over time.
2. **Input:** Input can be provided to or directly acquired by an AI system. The input is critical as it forms the basis for the AI system to generate outputs.
3. **Machine-based System:** At its core, an AI system is a machine-based system designed to operate with varying levels of autonomy and adaptiveness. It uses the input to infer outcomes, leveraging its model to process this information.
4. **Infers:** The central function of an AI system is to "infer". This means the system uses the input to reason and generate a different kind of output.
5. **Output:** An AI system produces output based on its inferences. The output can take the form of predictions, content, recommendations or decisions.
6. **Environment:** The environment, which can be physical or virtual, provides the context in which an AI system operates. The system's output can influence the environment and, in turn, the environment can provide new input for the system to process. In addition, the environment may impact the AI system's intended purpose.
7. **Autonomy and Adaptiveness:** These two capabilities are crucial components of an AI system. Autonomy allows the system to operate independently of human intervention, while adaptiveness is an optional characteristic, which refers to the system's ability to learn and improve over time based on new data and experiences.

Diagram I highlights the dynamic interaction of the system's objectives, inputs, inferences and outputs. It illustrates how an AI system can hold the capacity to continuously learn and adapt to enhance its functionality and impact on various environments.

The AI System definition given in the AI Act lacks elements that provide a comprehensive visual representation, making it challenging for various audiences to understand it easily. Therefore, our goal is to gradually introduce additional elements to the already established concepts. This will help in breaking down AI systems and adding to clarity.

Diagram II: Initially Expanded Visual Model of AI Systems under the AI Act



Source: Authors' own work and design following the AI Act and the OECD's definition of AI system

Diagram II introduces several additional elements:

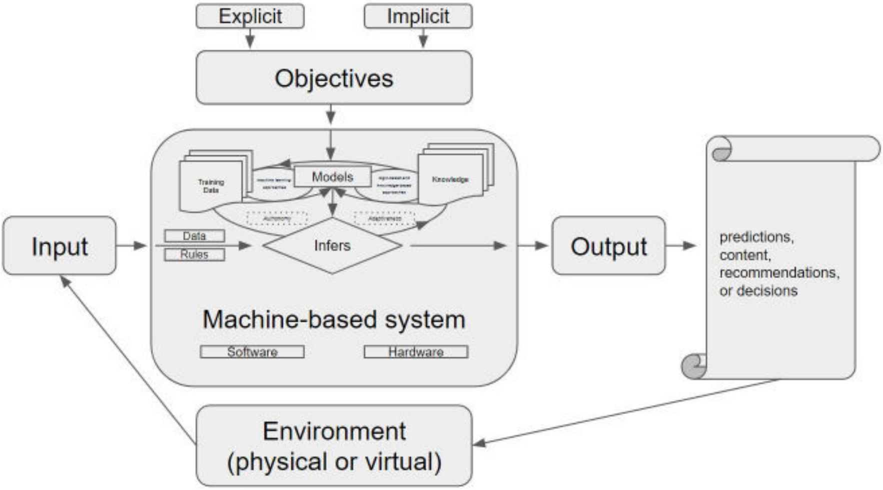
1. **Data** and **Rules** as inputs highlight diverse input types. Inputs can be provided by deployers, and users, or may be acquired from the environment.

2. **Model** forms the core of an AI system and represents all or part of the system's external environment (encompassing, e.g., processes, objects, ideas, people and/or interactions taking place in context) (OECD, 2022). Although Model is not specifically defined within the AI act, it is used extensively as a fundamental part of an AI system and included in the concept of "general-purpose AI model", which is defined separately. The recitals in the AI Act give explanations that along the AI value chain multiple parties often supply AI systems, tools and services, but also components or processes that are incorporated by the provider into an AI system with various objectives, including model training, model retraining, model testing and evaluation, integration into software, or other aspects of model development. A specific definition of "AI Model" is provided by the USA's Executive Order: "a component of an information system that implements AI technology and uses computational, statistical, or machine-learning techniques to produce outputs from a given set of inputs".

3. **Software** and **Hardware** components underscore the system's technological foundation.

These additions seek to enable a more detailed understanding of how AI systems function and process information to generate outputs.

Diagram III: Expanded Visual Model of AI Systems under the AI Act


Explicit
Implicit
Objectives
Training Data
machine learning approaches
Models
logic-based and knowledge-based approaches
Knowledge
Autonomy
Adaptiveness
Input
Data
Rules
Infers
Output
predictions, content, recommendations, or decisions
Machine-based system
Software
Hardware
Environment (physical or virtual)

Source: Authors' own work and design following the AI Act and the OECD's definition of AI system

1. **Training Data:** This component is essential for developing the models used by an AI system. Training data helps the system learn and improve its performance.
2. **Knowledge:** Represents the encoded information and expertise used by an AI system to draw inferences.
3. **AI Approaches:** This chiefly includes machine learning approaches and logic- and knowledge-based approaches:
    a. **Logic-driven/Symbolic AI:** Logic-driven AI, also known as symbolic or knowledge-based AI, works by creating formal models of phenomena, encoding them with symbols and structures, and applying logical reasoning processes to solve problems. An example is expert systems which can perform medical diagnoses using encoded knowledge from doctors or determine molecular structures of organic compounds. These systems are highly explainable and can be done with the same or greater accuracy than when done by humans, yet struggle with complex real-world scenarios.

    b. **Data-driven/Statistical AI:** This approach builds models from data through statistical methods and machine learning. Examples include neural networks and deep learning, which can manage complex tasks like image recognition by learning from large datasets.

While powerful, these systems usually lack explainability and require vast amounts of data to function effectively.

c. **Hybrid Systems:** Combining symbolic and sub-symbolic approaches, hybrid systems combine techniques to leverage the strengths of each. For example, NLP algorithms often combine statistical and symbolic methods (grammar rules).

These additions provide a comprehensive understanding of how AI systems function and how inferring is performed by showing the interplay within the model between training data, AI approaches and knowledge, stressing the AI system's autonomy and adaptiveness.

4 Key Characteristics of an AI System

The AI Act's definition of AI system and related recitals state that the key characteristics of an AI system that distinguish it from simpler traditional software systems or programming approaches lie in its capability to infer, but also the system's autonomy and adaptiveness. Although these key characteristics are not specifically defined in the regulation, they are described in the following manner:

- AI systems possess the capability to **infer**, which refers to the process of obtaining the outputs, such as predictions, content, recommendations or decisions, which can influence physical and virtual environments, and to the capability of AI systems to derive models or algorithms, or both, from inputs or data. An AI system's capacity to infer transcends basic data processing, and enables learning, reasoning or modelling.
- AI systems are designed to operate with varying levels of **autonomy**, meaning that they have some degree of independence of actions from human involvement and capabilities to operate without human intervention.
- AI systems can exhibit **adaptiveness** after being deployed, which refers to self-learning capabilities, allowing the system to change while in use.

5 Concepts of Inference, Autonomy, and Adaptability Across Different Fields

By analysing different definitions of the main concepts of an AI system – inference, autonomy and adaptability – we aim to explain their definition and application in different scientific fields, ranging from the social sciences through biology to information sciences.

”

Inference techniques in machine learning are used to make predictions, draw conclusions, or extract insights from data.

5.1 Inference

In logic, inference refers to the process of deriving conclusions from provided information or premises using acceptable reasoning methods. Several ways are commonly used to draw inferences: by deduction which, by analysing valid argument forms, draws out the conclusions implicit in their premises; by induction, which argues from many instances to a general statement; by probability, which passes from frequencies within a known domain to conclusions of stated likelihood; and by statistical reasoning, which concludes that, on average, a certain percentage of a set of entities will satisfy the stated conditions (Britannica, 2024). In that sense, inference forms part of reasoning, which makes reasoning a broader cognitive process of thinking and making conclusions. In other words, reasoning is the inferring of conclusions from premises (Over & Evans, 2024).

An ability to reason first emerged in our mammalian ancestors (Bennett, 2023), and according to numerous authors it is considered to be a characteristic reserved for some types of mammals, including humans. Inference and reasoning are uniquely human abilities that set us apart from traditional computers. Inference in machine learning is the process where a trained model is used to make predictions or classifications on new, unseen data (Anderson, 2023). Inference techniques in machine learning are used to make predictions, draw conclusions, or extract insights from data (Parti, 2024). However, the weak reasoning capabilities of, for example, large language models (LLMs) are partially compensated by their extensive associative memory. What may appear to be commonsense reasoning in these models is often more akin to pattern matching achieved through the analysis of vast amounts of text (Bennett, 2023).

For example, by training ChatGPT-4 to predict not just the final output in the form of an answer to the inquiry, but also the subsequent steps in the reasoning process, the model starts to demonstrate emergent properties of thinking. It is nevertheless important to note that this is not true thinking in the human sense; instead, it is more like creating a simulation of thought processes (Bennett, 2023).

5.2 Autonomy

In Western ethics and political philosophy, autonomy is described as the state or condition of self-governance, or leading one's life according to reasons, values or desires that are authentically one's own (Taylor & Stacey, 2017).

In sociology, autonomy is a relative concept, shaped by an individual's social position, resources for pursuing an autonomous life, prevailing norms of autonomy, and perceived constraints and demands (Börner et al., 2020). This concept has evolved, with three distinct modes of autonomy corresponding to different stages of modernity: (a) the normative idea of autonomy prevalent in the early phases of modernity; (b) individual autonomy claims at the micro level; and (c) institutionalised demands for autonomy that have gained prominence in more recent times (Börner et al., 2020). The theme of political autonomy has its roots in the concept of autonomia in Ancient Greek cities, while individual autonomy emerged alongside the Protestant ethic and humanist thought, as illustrated in Kant's notion of the "autonomy of reason" (Sapiro, 2019).

In sociological analyses, autonomy plays a critical role in examining the production and circulation of symbolic goods, exploring their relationships with economic, social and political conditions without merely reducing these goods to those conditions. Yet, interpretations of autonomy vary across three research strands that have engaged with this concept since the 1950s–1960s: the sociology of professions, Marxist reflection theory, and field theory (Sapiro, 2019). In the sociology of professions, autonomy is more closely aligned with its political implications, referring to the organisational structure of professions where their 'technical' and organisational independence – concerning access control and practices – is recognised by the state. In contrast, the Marxist tradition positions autonomy as a critique of theories that regard artworks, like religion and the state, as a superstructure reflecting social production relations. Conversely, in field theory, autonomy pertains to the relative independence of production spheres from external political, religious or economic pressures (Sapiro, 2019).

In biology, the concept of autonomy pertains to a system's ability to define and uphold its own rules and behaviours actively. This idea is critical for understanding living systems on an individual level, highlighting two primary aspects: (1) self-construction, where organisms continuously build and maintain the components essential for their functioning through cellular processes; and (2) functional interaction with the environment, whereby organisms act as agents that adapt their surroundings to sustain their existence as non-equilibrium systems. Biological autonomy thus portrays organisms as organised entities capable of self-production, self-maintenance, goal-setting, norm-establishment, and environmental interaction to ensure their continued existence and development (Moreno, 2013; Moreno & Mossio, 2015).

Machine autonomy can be understood as "the ability of a computer to follow a complex algorithm in response to environmental inputs, independently of real-time human input" (Etzioni & Etzioni, 2016 according to Formosa, 2021). Autonomy in the sense of autonomous systems means that systems can generate their control mechanisms and employ these mechanisms to sustain themselves. A system achieves autonomy when the internal organisation of its processes becomes crucial for its self-preservation, assuring both its functionality and that of the processes that support its autonomy (Collier, 2002).

Finally, in AI research autonomy refers to systems that function independently of human control. Within technical discourse, autonomy is also linked to the capacity of AI systems to learn and make decisions based on their own experience (Prunkl, 2021).

5.3 Adaptability

The word "adaptation" does not stem from its current use in evolutionary biology, but dates back to the early 17th century when it indicated a relationship between design and function and referred to how something fits into something else (Gittleman, 2022).

In Talcott Parsons' AGIL schema, adaptation is one of the four essential functional imperatives that any social system must address to maintain stability and achieve its goals (Parsons, 2013). The locus of adaptation is the "physico-chemical system" and its prime function is to conceptualise the human organism's exchanges with the natural world (Chernilo, 2017). This process results from the interplay of a society's value patterns, normative structures, and goal capacities with its environment's instrumental and efficient pressures. The outcome, termed a society's "adaptive structure", signifies a social dimension that imposes abstract pressures, limiting the possibilities for social arrangements (Alexander, 2014).

Piaget (1932) defined adaptation as the equilibrium toward which an organism progresses through its interactions with the environment. He posited that adaptation involves moving away from a wholly egocentric perspective and developing internalised operations (Motamedi, 1977). Further, adaptability encompasses a social system's capacity to modify either the external environment, its internal conditions to meet external demands, or both simultaneously to achieve a fit (Motamedi, 1977). This entails sensing and comprehending both internal and external environments – the total environment – and taking action to create alignment between the two. Social systems can employ at least four groups of tactics to facilitate adaptation in specific circumstances (Motamedi, 1977).

In biology, adaptation is the process by which a species becomes fitted to its environment; it is the result of natural selection acting upon heritable variation over several generations (Gittleman, 2022). In addition, the term "adaptation" commonly refers to the process of becoming adapted or to the features of organisms that promote reproductive success relative to other possible features (Gittleman, 2022). Adaptation relates to the process by which organisms respond to new environmental conditions. It involves immediate survival responses that may not initially be inherited but could eventually become heritable via genetic changes like mutations or genomic reorganisation. The evolution of adaptive responses requires modifications in regulatory and developmental networks, which become increasingly complex as the number of components involved grows. However, adaptability in individuals serves as both a buffer and a catalyst for evolutionary change, allowing for gradual adjustments that contribute to the overall adaptive capability of a species over time (Bateson, 2017).

> The evolution of adaptive responses requires modifications in regulatory and developmental networks, which become increasingly complex as the number of components involved increases.

Adaptability is the extent to which a software system adapts to change in its environment. An adaptable software system can tolerate changes in its environment without external intervention (Subramanian & Chung, 2001). Adaptive AI systems are designed to enhance decision-making by prioritising swift responses and maintaining flexibility to accommodate unforeseen challenges. These systems focus on continual learning from real-time data to swiftly adjust to evolving real-world conditions (Gartner Glossary, 2024).

5.4 Should AI Be Inferential, Autonomous, and Adaptable?

We analysed the concepts of inferring, autonomy and adaptability in order to try to understand their differences and the importance of the existence and implementation of these concepts during the development and application of AI systems. Analysing these concepts through the prism of different scientific fields, we came to the conclusion that the development of AI systems which include the mentioned concepts is of great importance. Namely, by using inference with AI systems, we ensure the creation of AI systems that can draw conclusions from training models with greater certainty and precision, make predictions, and increasingly resemble human cognitive functions. Although very similar to the sociological and biological definition of autonomy, the autonomy of AI systems should assure an increase in the efficiency of the systems themselves, their independence, coupled with an increase in human efficiency and reduction in the time required for controlling them. The concept of adaptability, which we similarly apply in the social and natural sciences, implies that no organism or social system, including an AI system, can survive for long without the indispensable need to adapt to external and internal factors that must influence its evolution and development. All in all, the concepts explored are vital for the development of effective AI systems.

6 What Is (Not) an AI System?

Differentiating AI systems from other IT products, such as regular software, computer systems, and smart systems, is essential for proper regulation and application. Pursuant to the AI Act, the main characteristics that define AI systems include inference autonomy, and sometimes adaptability. AI systems are designed to operate with varying levels of autonomy and can infer, meaning they can reason and generate outputs based on input data. These systems use models, which can be developed mostly through machine learning or logic-based approaches, to process input data and achieve their objectives. Unlike traditional software that follows predefined rules, AI systems can possess the capacity to learn and adapt, boosting their functionality over time.

Conversely, systems that do not have these characteristics are probably not considered AI systems. Simple automated systems that operate based on static rules without the ability to learn or adapt, manual processes requiring human intervention, and conventional software algorithms that lack learning capabilities fall outside the scope of AI as defined by the AI Act. For instance, a spreadsheet using predefined formulas, as well as SPSS, which processes data without inferring, should not be considered as AI. Understanding these distinctions helps organisations comply with regulations and determine whether their systems fall under the AI Act's provisions. This definition, based on characteristics rather than techniques, ensures clear differentiation and adaptability for future regulations (Trincado Castán, 2024).

Automated decision-making and profiling are two distinct types of personal data processing highlighted by the GDPR, each defined with unique parameters (Richardson, 2022; Krivokapić, Nikolić, 2023). Automated decision-making refers to decisions made without human intervention, primarily based on algorithmic processes. In contrast, profiling involves the use of personal data to evaluate aspects of an individual, such as predicting their behaviour, preferences or abilities. Distinguishing these two processes is crucial because they raise different challenges and considerations for data protection. By differentiating them, the GDPR underscores the need for transparency, fairness and accountability in both areas, making sure that individuals' rights are protected in the rapidly evolving digital landscape.

Profiling can be performed without using AI. While AI and machine learning are regarded as advanced tools for profiling, traditional statistical methods, rule-based systems and simpler algorithms can

> Automated decision-making and profiling are two distinct types of personal data processing highlighted by the GDPR.

also be employed to create profiles. For example, credit scoring systems were using predefined rules and criteria to assess individuals' creditworthiness long before the emergence of AI-driven systems. Similarly, automated decision-making can operate without profiling or AI technologies. For instance, consider a basic online form that automatically approves or denies an application based on user-entered criteria. If an applicant inputs an age below a specified threshold, the system might automatically deny the application. This type of automated decision-making does not require profiling or AI technologies. Before the widespread adoption of AI, many automated systems relied on rule-based algorithms and straightforward criteria for decision-making.

> The core principles of the GDPR and the AI Act are closely intertwined, leading to a significant overlap in the requirements for trustworthy AI and the established GDPR principles.

The core principles of the GDPR and the AI Act are closely intertwined, leading to a significant overlap in the requirements for trustworthy AI and the established GDPR principles. Despite the GDPR not explicitly mentioning AI, it includes provisions critical for personal data processing by AI systems, such as automated decision-making and profiling. These cases trigger specific obligations like the right to information and the right to object. Even if AI is not involved in automated decision-making or profiling, all processing of personal data must comply with relevant data protection regulations and principles to safeguard individuals' privacy rights. Controllers must carefully navigate the tension between traditional data protection principles – such as purpose limitation, data minimisation, sensitive data categorisation, and restrictions on automated decisions – and the expansive capabilities of AI. However, it is possible to interpret and adapt data protection principles to align with beneficial AI applications. Requirements applicable to automated decision-making or profiling should thus be considered, even if these processes are not explicitly recognised in specific cases.

7 Conclusion

It was very important to determine a clear definition of AI system able to ensure legal certainty by providing a concrete, yet neutral definition covering specific AI systems and all potential technological developments. The European Commission's initial proposed definition of AI faced criticism for being too broad. After a complex negotiation process, where all aspects of the definition of AI system were scrutinised, including the different techniques that could be used, the level of an AI system's autonomy, and its influence on physical and virtual environments, the final definition adopted in the AI Act was inspired by the OECD definition, which is broadly accepted. Understanding of the prescribed AI system definition is challenging and hence we developed visual models so as to foster a deeper, more accessible understanding of AI systems and its core elements under the AI Act. The presented visual models serve as a practical tool for demystifying these complexities. Their task is to provide a clear, graphical representation of how AI systems operate, including an illustration of the inputs, processes and outputs.

Finally, it is essential to stress the importance of having a uniform AI system definition. Otherwise, each legal regulation could apply different or inconsistent definitions of an AI system and, in turn, certain systems using the same technical approaches or possessing the same characteristics could fall within the scope of one legal regulation, but not another. For example, in case different legal regulations have their own definition of AI system, there is a risk that the same system could be considered as an AI system entailing high risk under the AI Act and could not even be considered as AI according to the rules of other regulations.

REFERENCES

- Alexander, J. (2014). Modern Reconstruction of Classical Thought: Talcott Parsons. Routledge.
- Anderson, K. (2023). What is machine learning inference? Retrieved 10 July 2024 from: https://telnyx.com/resources/machine-learning-inference
- Bateson, P. (2017). Adaptability and evolution. Interface Focus, 7(5), p. 20160126. doi: https://doi.org/10.1098/rsfs.2016.0126.
- Bennett, M. S. (2023). A Brief History of Intelligence: Evolution, AI, and the Five Breakthroughs that Made Our Brains. Mariner Books.
- Börner, S., Petersen, N., Rosa, H., Stiegler, A. (2020). Paradoxes of late-modern autonomy imperatives: Reconciling individual claims and institutional demands in everyday practice. The British Journal of Sociology, 71(2), 236–252.
- Britannica, T. Editors of Encyclopaedia. (2024). inference. Encyclopedia Britannica. Retrieved 10 July 2024 from: https://www.britannica.com/topic/inference-reason
- Castán, C. T. (2024). The legal concept of artificial intelligence: the debate surrounding the definition of AI System in the AI Act. BioLaw Journal-Rivista di BioDiritto, (1), 305-344. Retrieved 15 July 2024 from: https://teseo.unitn.it/biolaw/article/view/3000
- Chernilo, D. (2017). Adaptation: Talcott Parsons. In Debating Humanity: Towards a Philosophical Sociology , 87–110. Cambridge: Cambridge University Press.
- Collier, J. (2002). What is Autonomy? Retrieved 10 July 2024 from: https://web-archive.southampton.ac.uk/cogprints.org/2289/3/autonomy.pdf.
- Formosa, P. (2021). Robot Autonomy vs. Human Autonomy: Social Robots, Artificial Intelligence (AI), and the Nature of Autonomy. Minds and Machines, 31(4), 595–616.
- Garner Glossary. (2024). Definition of adaptive AI. Retrieved 10 July 2024 from: https://www.gartner.com/en/information-technology/glossary/adaptive-ai#:~:text=Adaptive%20AI%20systems%20support%20a,changes%20in%20real%2Dworld%20circumstances.
- Gittleman, J. L. (2022). adaptation. Encyclopedia Britannica. Retrieved 10 July 2024 from: https://www.britannica.com/science/adaptation-biology-and-physiology
- Krivokapic. Đ., Nikolic, A. (2023). Application of Data Protection Standards to Artificial Intelligence Systems: Automated Decision Making and Profiling. Kopaonik School of Natural Law – Slobodan Perović, Vol. III, 151-174.
- Marwala, T. (2017). Rational choice and artificial intelligence. arXiv preprint arXiv:1703.10098.
- Moreno, A. (2013). Autonomy. In: Dubitzky, W., Wolkenhauer, O., Cho, KH., Yokota, H. (eds) Encyclopedia of Systems Biology. Springer.
- Mossio, M., & Moreno, A. (2015). Biological autonomy: A philosophical and theoretical enquiry. History, philosophy, and theory of life sciences series. Springer.
- Motamedi, K. K. (1977). Adaptability and capability: A study of social systems, their environment, and survival. Group & Organization Studies, 2(4), 480-490.

- OECD. (2022). Framework for the Classification of AI systems. Retrieved 15 July 2024 from: https://www.oecd-ilibrary.org/science-and-technology/oecd-framework-for-the-classification-of-ai-systems_cb6d9eca-en
- OECD. (2024). Recommendation of the Council on Artificial Intelligence. Retrieved 15 July 2024 from: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449
- Over, D. E., & Evans, J. S. B. T. (2024). Human Reasoning. Cambridge: Cambridge University Press.
- Parsons, T. (2013). The social system. Routledge.
- Parti, A. (2024). Machine Learning Inference - All You Need to Know. Retrieved 10 July 2024 from: https://pareto.ai/blog/machine-learning-inference
- Prunkl, C. (2021). Is there a trade-off between human autonomy and the 'autonomy' of AI systems? In Conference on Philosophy and Theory of Artificial Intelligence, 67-71. Cham: Springer International Publishing.
- Ruschemeier, H. (2023). AI as a challenge for legal regulation–the scope of application of the artificial intelligence act proposal. In Era Forum, Vol. 23, No. 3, 361-376.
- Russell, S. J., & Norvig, P. (2020). Artificial intelligence: a modern approach. Upper Saddle River.
- Sapiro, G. (2019). Rethinking the concept of autonomy for the sociology of symbolic goods. Biens Symboliques/Symbolic Goods. Revue de sciences sociales sur les arts, la culture et les idées, (4).
- Sarker, I. H. (2022). AI-based modeling: techniques, applications and research issues towards automation, intelligent and smart systems. SN Computer Science, 3(2), 158.
- Subramanian, N., & Chung, L. (2001). Metrics for software adaptability. Proc. Software Quality Management (SQM 2001), 158.
- Taylor, J. S. (2017). autonomy. Encyclopedia Britannica. Retrieved 10 July 2024 from: https://www.britannica.com/topic/autonomy
- The White House. (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Retrieved 5 July 2024 from: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-se-cure-and-trustworthy-development-and-use-of-artificial-intelligence/

Chapter 2

Beyond the Hype: Navigating AI Regulation Through the GDPR and the AI Act in France and Europe

Sébastien Fassiaux

1 Introduction

Artificial intelligence (AI) is sometimes perceived as a new, unregulated technology. This misconception is largely due to generative AI making the headlines following the introduction of ChatGPT in late 2022. The new chatbot created by OpenAI is one of the fastest growing consumer products of all time, with 1 million users after 1 week and 100 million within 2 months (Milmo, 2023). Its release has clearly marked an inflexion point in consumers' awareness of AI's potential risks and opportunities. Not only has ChatGPT impressed users with its ability to answer most queries in any language and perform tasks almost instantly, but it has also captured the imagination of market actors for developing applications based on its capabilities. As a result, the EU legislator introduced provisions related to general-purpose AI models in the newly enacted Artificial Intelligence Act, giving the impression that these new rules were the first ones applicable to AI systems.

Still, AI systems are not created and implemented in a legal void. In fact, before the hype surrounding its generative form, AI was long operating in the background on the Internet – powering search engines, personalised advertising, content recommendations, maps directions etc. – without consumers necessarily realising that AI was involved. Due to their dependency on data – including personal data, AI systems have been subject to data protection laws for some time now, even prior to the entry into force of the EU's General Data Protection Regulation (GDPR). Today, arguably the most important limits on the use of AI systems are set by data protection and competition authorities.

While generative AI has now brought AI to the forefront of public debate, the EU was already working on enacting the AI Act before ChatGPT's release. In this context, exploring the interactions between the AI Act and the GDPR is relevant because the former will be fully applicable only in 2027, and its effects will take even longer to materialise. More generally, a better understanding of the overlap between AI and privacy policy issues appears essential for the future of AI development (OECD, 2024).

This contribution thus highlights the extent to which data protection regulation already regulates the use of AI systems. It also explores how competition and copyright laws – among other fields – further complement privacy rules in AI regulation. It discusses the extent to which the AI Act brings novelties to further regulate AI systems developed and deployed by private and public entities and examines the interactions between the AI Act and the GDPR. The analysis focuses on the case of France, where the data protection and competition authorities have been handling important cases concerning AI systems mainly in the fields of facial recognition, digital advertising, and generative AI, fining companies such as Clearview AI, Google and Meta.

> ” While generative AI has now brought AI to the forefront of public debate, the EU was already working on enacting the AI Act before ChatGPT's release.

The paper also explores how the French data protection authority (CNIL) cooperates with other public and private entities in view of regulating AI systems. For instance, it discusses how the CNIL collaborates with other EU regulators and within the European Data Protection Board's taskforce dedicated to the coordination of enforcement efforts related to generative AI systems.

Finally, the paper examines the CNIL's launch of regulatory sandboxes designed to assist innovative projects – including those using AI systems – with GDPR compliance, and the new role the authority will have in this space under the AI Act.

2 Background on AI Regulation

The fact that new legislation aiming to regulate the use of AI systems is being adopted on the EU level does not mean that the law previously disregarded the issue. Tech companies have been using AI systems in consumer markets since the early days of the Internet. To a large extent, AI systems powering search engines, personalised advertising, and recommender systems made the fortunes of Google and Facebook with their advertising businesses. These companies are now among the most valuable on the planet. Yet, their use of AI systems for making huge quantities of information readily available and increasingly personalised has not been left unchecked. While the focus has mostly been on the regulation of AI through data protection law (Section 2.1), competition and intellectual property law are becoming increasingly relevant (Section 2.2).

2.1 AI Regulation under Data Protection Law

Data protection law was initially used to regulate the matter. This is not surprising given the dependency of AI systems on data for their training and operations. Important cases actually predate the GDPR and relate to fundamental questions over the balance between the right to privacy, the right to the commercial exploitation of personal data, and the freedom of expression and information. The issue of de-referencing is emblematic of this balance.

In the landmark Google Spain case of 2014, the Court of Justice of the EU (CJEU) established the so-called 'right to be forgotten' (case C-131/12). This case raised the issue of an individual's right to be removed from a search engine's results based on the interpretation of the 1995 Data Protection Directive (95/46/EC), read in light of the fundamental rights to privacy and to the protection of personal data guaranteed by the Charter of Fundamental Rights of the EU. Ultimately, the CJEU prioritised the fundamental right to privacy over the economic interests of the operator of an AI-powered search engine and the general public's interest in having access to personal information, subject to safeguards when the relevant person plays a role in public life. The judgment also emphasised the CJEU's role as a prominent guardian of individuals' privacy in Europe and the right it established was codified in the GDPR a few years later.

This case – and many others – show that important legal questions concerning AI have been in the open for some time now and did not appear only in the advent of the GDPR or – a fortiori – the AI Act. Indeed, the plaintiff in the Google Spain case had complained to the Spanish data protection authority in 2010, almost 15 years ago and 6 years before the GDPR was adopted.

Since Google Spain, the CJEU has continued to adjudicate important questions related to AI systems based on data protection law. For example, still with regard to Google's search engine, in 2019 the CJEU also established a "right to be accurately remembered" (European Commission, Legal service, 2022: 119), based on the GDPR read in light of the Charter of Fundamental Rights of the EU (case C-136/17). Going further than pure removal, this right can compel search engines to modify the ranking of their searches, directly affecting their search algorithms.

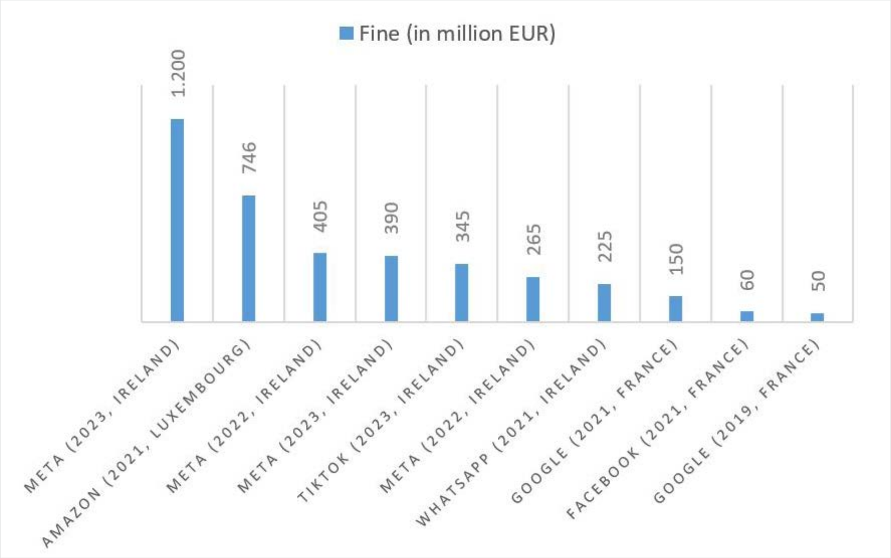
Also in 2019, the CJEU delimited the territorial scope of the right to be removed from a search engine's results (case C-507/17). The case originated in France where the CNIL had imposed a fine of EUR 100,000 on Google for refusing to remove results involving personal data to all worldwide versions of its search engine. The CJEU found that when a search engine like Google grants a request for removal, it is in principle required to carry out the removal only with respect to versions available in the EU, not elsewhere. The French Conseil d'Etat, which had referred the question to the CJEU, finally annulled the CNIL's fine on Google in 2020 (decision n°399922).

As these pre-GDPR cases show, data protection requirements regarding AI systems thus preceded the recent adoption of the AI Act, the Digital Services Act (DSA), or the Digital Markets Act (DMA), which have established new rules and standards in this space.

At any rate, the tension between the fundamental right to privacy and the commercial interests of large companies using AI systems is very much alive today, 6 years after the GDPR started to apply. Although early GDPR enforcement has been slow, its bite is now growing.

Figure 1 below shows the biggest fines issued under the GDPR by June 2024. All of these fines were imposed on companies heavily relying on huge quantities of personal data for their AI systems to drive personalised advertising, as well as content and product recommendations, among others. Overall, the effectiveness of those fines and accompanying injunctions to deter tech firms from engaging in GDPR infringements is up for discussion. Still, their general influence on industry practices with regard to the handling of personal data elevates data protection law to the forefront of AI regulation.

Figure 1: Highest fines imposed under the GDPR by June 2024



Source: DPC, CNPD, CNIL

Despite all of the affected companies having their EU establishments in Ireland and Luxembourg – thus giving data protection authorities there a preponderant influence in enforcing the GDPR due to its one-stop-shop mechanism, the graph shows the active role of the CNIL, which imposed three of the ten biggest fines under the GDPR. The CNIL's enforcement action is to be further discussed in Section 4 below.

2.2 AI Regulation under Competition and Intellectual Property Law

Although the focus of this contribution is on exploring the interplay between the GDPR and the AI Act, it is also relevant to note that AI regulation is increasingly taking place in other fields, including competition and intellectual property law. The scope and objectives of all three areas differ substantially, though.

The main objective of data protection law is to protect individuals' privacy and its scope is quite broad, as exemplified by the AI-related cases in France discussed in Section 4. Being technology-neutral and applying horizontally to both private and public entities, the GDPR applies to a wide range of practices involving AI systems. Its comprehensive rules and principles leave sufficient room for interpretation, which is both its strength and weakness.

Instead, the main objective of competition law in this context is to ensure that AI-intensive markets remain fair and contestable, guaranteeing a level playing field for businesses, ultimately benefitting consumers. Here, trustbusters have to strike a difficult balance between innovation and competition on the merits. Traditional competition enforcement based on Articles 101 and 102 of the Treaty on the Functioning of the EU has led to important fines, also in AI-related markets. For instance, the largest-ever competition fine of EUR 4.34 billion was imposed by the European Commission on Google in 2018 – it was later reduced to EUR 4.125 billion by the CJEU. The Commission found that Google had abused its dominance in the market for mobile operating systems with Android to safeguard its dominance on the market for general search services, thereby protecting its main source of revenue from its AI-driven search engine. Yet, traditional, ex post competition rules have been difficult to enforce in digital markets. Proceedings take years to result in sanctions and, in the meantime, the harm to competitors and consumers is irreversible.

In order to reverse this trend, the EU legislator adopted the Digital Markets Act (DMA) in 2022. Designed as an ex-ante regulation mainly concerned with ensuring competitive markets, the DMA complements traditional competition rules by applying more specific obligations in a number of defined digital markets controlled by gatekeepers designated as such by the Commission. While DMA enforcement is still in its infancy, some experts already argue that, contrary to the AI Act, the DMA contains "the most far-reaching, most overlooked but potentially also most effective regulatory constraints for AI" (Hacker, Cordes & Rochon, 2024: 51). At a recent workshop on generative AI, EU Commissioner for Competition Margrethe Vestager also confirmed that the DMA applies to generative AI features embedded in services covered by the regulation (European Commission, 2024):

"We continue to apply our trusty merger and antitrust rules. Even though we sometimes have to remind AI market players that the competition rules also apply to them. And our Digital Markets Act applies too: the DMA can also regulate AI even though it is not listed as a core platform service itself. AI is covered where it is embedded in designated core platform services such as search engines, operating systems and social networking services. So we are applying our rulebook to concerns that we have already in the AI world. We are looking at the issues very closely, from all angles and with all our tools."

What is sure is that the DMA will profoundly affect AI-related markets and provide enforcers with additional tools to regulate them. For instance, after the German competition authority had innovated in 2019 by considering personal data handling by dominant firms (finding Meta had inappropriately combined personal data across its AI-driven services), the DMA now also considers such conduct by gatekeepers.

At the same time, discussions around copyright are gaining importance in the context of the development of generative AI applications and their underlying foundation models. Indeed, such models require training on very large datasets, but developers have been accused of training them using copyrighted materials. In this respect, the 2019 Directive on Copyright in the Digital Single Market (2019/790, or CDSM) allows reproductions and extractions for text and data mining in specific conditions. Rightsholders can reserve their rights to prevent such mining unless it is for scientific research. If rights are reserved, AI model providers must obtain authorisation from rightsholders for text and data mining. This possibility for rightsholders to opt-out from the exemption of copyright protection has been criticised for going precisely beyond copyright protection. Instead, by allocating property rights for AI's "building blocks" the EU legislator appears to have adopted a "property-right approach to the regulation of AI" (Margoni & Kretschmer, 2022: 688). What is at stake is the tension between the ability to innovate without authorisations and the protection of authors. The AI Act follows this approach by requiring providers of general-purpose AI models to put in place a policy to comply with EU copyright rules, including with the text and data mining exemption under the CDSM.

This is particularly relevant for the news media, which first had to adapt to the development of the Internet, then to digital platforms, and now see a threat from generative AI applications. In fact, the French competition authority imposed a fine of EUR 250 million on Google in March 2024 for breaching commitments in a case where the tech company was found to have infringed the French law transposing the CDSM and aimed at ensuring fair negotiations between online platforms, press agencies, and publishers (decision 24-D-03). The regulator also found that Google's AI service Bard (now renamed Gemini) utilised content from press agencies and publishers without notification or opt-out options, hindering their ability to negotiate fair remuneration. This fine represents more than what the CNIL has ever imposed in a single year for GDPR infringements. This case underlines how competition and copyright issues are equally relevant for AI regulation.

3 The AI Act and Its Interplay with the GDPR

The AI Act integrates with the strategy outlined in the European Commission's 2020 White Paper on Artificial Intelligence by operationalising its vision of an ecosystem of excellence and trust (European Commission, 2020a). The White Paper stressed the need to establish a regulatory framework addressing high-risk AI systems by setting requirements for data governance, transparency, and cybersecurity, involving both pre-market evaluations and post-market monitoring. In line with the objective of fostering trustworthy AI, the Commission highlighted the need to protect the fundamental right to privacy in the context of AI regulation. The EU's AI strategy is ultimately furthering the goals of its digital strategy, outlining policy measures to advance the digital transition with regard to skills, infrastructures, public services and the economy (European Commission, 2020b & 2021).

As a result, the Commission prepared a proposal for an AI Act in 2021. The text was finally approved by the Council in May 2024 and entered into force on 1 August 2024. The AI Act is the first comprehensive set of rules specifically applying to AI systems in the world. As such, the Commission aims to promote it as a global standard for AI systems regulation. These rules are the outcome of intense negotiations among the Commission, Parliament and Council. With around 3,000 amendments, it is probably the most debated legislative file of the previous legislature. Difficult negotiations took place especially as concerns foundational models. France partially opposed the regulation to foster innovation and the development of European AI models aligned with local language and culture. Representing the views of liberal leaders, Jean-Noël Barrot explained as French Minister for Digital Transition in November 2023 that there was still hope for European models to develop in the coming years, although that regulation should not hinder innovation and instead focus on protecting consumers and citizens (CNIL, 2024: 8). Ultimately, a relevant question is whether the AI Act will effectively become a global standard just like the GDPR for privacy, or if its Brussels effect will be limited (Bradford, 2021).

The Act complements the data protection, competition, and copyright law regimes applicable to AI systems. It is mainly a 'products' regulation and seeks to ensure that AI systems placed on the market are safe – just like any other products – and respect fundamental rights. It does so by categorising AI systems by risk level and imposes obligations accordingly. In practice, this means that most AI systems will not be heavily regulated.

The regulation requires CE marking for high-risk uses, achieved through the compliance with significant obligations for AI systems providers, deployers, importers and distributors, controlled by rigorous audits. General-purpose AI models are generally subject to limited obligations, for instance related to documenting their sources of training, and respect for copyright, except those involving systemic risks based on their increased capabilities, which are also subject to considerable obligations. But low-risk AI systems face lighter requirements, for example in terms of transparency. Yet, in order to protect fundamental rights and European values, some AI systems involving unacceptable risks are prohibited outright. For instance, AI systems used for social scoring or to infer emotions in the workplace or educational institutions are prohibited.

The AI Act interacts with the GDPR in several ways. The scope and objectives of both instruments have already been discussed. Importantly, the AI Act specified that it does not alter existing EU laws on personal data processing or the duties of supervisory authorities overseeing compliance. It thus maintains the obligations of AI providers and deployers as data controllers or processors under EU or national data protection laws. Similarly, data subjects retain all rights under these laws, including those related to automated decision-making and profiling. Two limited exceptions nonetheless exist and refer to data quality and the competence of national data protection authorities. Another one is discussed in Section 6 and relates to regulatory sandboxes.

As such, the AI Act complements the GDPR because, unlike the latter, it is not only concerned about personal data. A good example is the requirements for data quality. Indeed, the Act mandates high-quality data to ensure that AI systems are safe, effective, unbiased and non-discriminatory, adhering to data protection laws like the GDPR. Data used for AI training and testing must be relevant, representative, complete, and error-free, with clear transparency concerning data collection purposes. Privacy-preserving techniques should be used, and data should reflect specific usage contexts. Still, these requirements are not only limited to personal data.

Yet, the AI Act foresees an exception to the strict GDPR rules on special categories of personal data, i.e., personal data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data for unique identification, data concerning health, and data regarding a natural person's sex life or sexual orientation. The GDPR indeed in principle prohibits the processing of such data and only allows it in limited circumstances. However, the AI Act add a new exception by giving the possibility for providers of high-risk AI systems to exceptionally process these sensitive data to detect and correct bias in their datasets, if strictly necessary, with their use being subject to strict conditions. Despite being an exception to the prohibition of sensitive data processing, this possibility should be welcomed to guarantee that AI systems are trained and tested on unbiased datasets, maximising the respect of fundamental rights when deployed.

Another way that the AI Act complements the GDPR is by extending the competence of national data protection authorities. As anticipated above, the AI Act puts in place a system of post-market monitoring. Under this system, national market surveillance authorities are tasked with controlling compliance with the AI Act, except for general-purpose AI models for which the regulation grants exclusive competence to the European Commission. Its newly established AI Office will enforce rules and monitor compliance. Of course, member states already have market surveillance authorities to monitor compliance with product safety rules. Yet the Act specifies that for high-risk AI systems used in law enforcement, border control management, and the administration of justice and democratic processes – including the use of biometrics in these three areas – member states need to designate their supervisory authorities under either the GDPR or the Law Enforcement Directive (2016/680) for market surveillance and control. In France, the CNIL, which is already the data protection authority under both instruments, will thus have extended competences under the AI Act. This might create additional bottlenecks for national data protection authorities which already suffer from a lack of resources, a challenge recently identified by the EU Agency for Fundamental Rights (2024).

4 Regulating AI Through Data Protection Enforcement in France

Given the decentralised nature of the enforcement system under the GDPR, each member state has designated a supervisory authority to enforce it. France designated the *Commission nationale de l'informatique et des libertés* (CNIL), established in 1978 as an independent administrative body to act as the country's data protection authority. Under the GDPR, it can now impose fines of up to EUR 20 million or, for companies, up to 4% of their total worldwide annual turnover, whichever is higher.

For some years now – and before the hype around generative AI, the CNIL has reflected on AI development and regulation. Already in 2017, it published a report following a public debate on the ethical challenges of AI (CNIL, 2017). In its 2018 annual report – the year the GDPR began to apply, the CNIL dedicated a section to AI, chiefly focusing on its ethical and societal implications (CNIL, 2019). It emphasised the importance of educating the public about AI's potential and fostering a sustainable AI model on the national, European and international levels. Moreover, it underscored the need for ethical considerations in AI, addressing issues like human autonomy, algorithmic discrimination, and the impact of hyper-personalisation on societal structures. Given the importance of personal data for training and operating AI systems, it also stressed the GDPR's regulatory role in creating a trustworthy framework for data use in AI, ensuring transparency and individual rights.

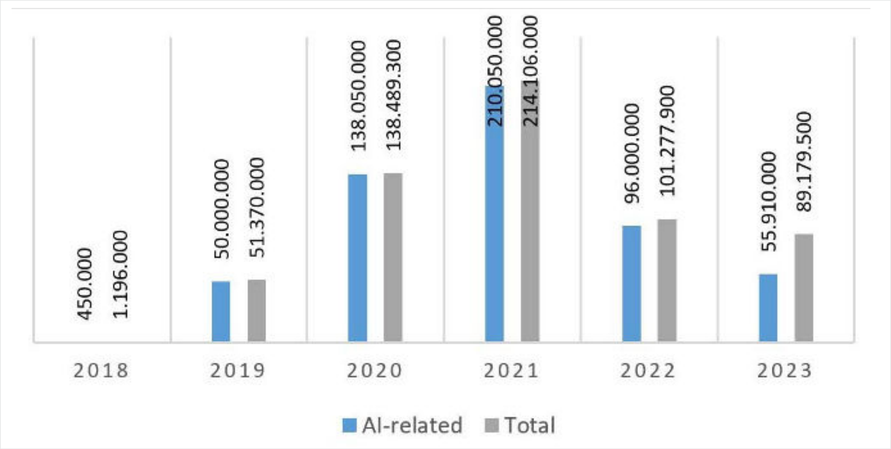
In this context, the CNIL has long addressed AI issues through case studies, such as the use of smart cameras in public spaces, voice assistants, and facial recognition technology. It has also reviewed government projects like the AI tools for tax fraud detection. Since 2022, the CNIL has published resources clarifying AI-related challenges and GDPR compliance, offering an AI self-assessment guide for the public and professionals.

At a CNIL event on AI and free will held in November 2024, its president Marie-Laure Denis explained that, while a pause in AI innovation was not the issue, one could not wait for the AI Act to regulate AI systems and that the GDPR was sufficiently flexible to assure positive AI development. As developed below, the CNIL indeed has a proven track record of active GDPR enforcement in different AI fields (Section 4.1) and has already started working on its application to generative AI, along with its European peers (Section 4.2).

4.1 Early GDPR Enforcement

The CNIL has been one of the most active GDPR enforcers in Europe, also in the AI field. In fact, the GDPR marked a breaking point in terms of sanctioning the mishandling of personal data (see Figure 2 below). Pre-GDPR, the French regulator had been imposing modest fines. In 2016, the CNIL imposed fines worth a total amount of EUR 160,000, and more than doubled them in 2017 (EUR 371,000). Post-GDPR, that is after the regulation became applicable in May 2018, the regulator's fines reached record-high levels, with individual sanctions in the millions, even reaching EUR 150 million for a fine imposed on Google in 2021.

Figure 2: Share of CNIL fines (in EUR) on entities using AI systems compared to total fines imposed (2018–2023)



Source: CNIL

Interestingly, a closer look at the nature of the data processing activities at hand shows that fines particularly affected entities heavily relying on personal data for their AI-powered services. In fact, the overwhelming majority of fines were imposed on companies for their AI-related activities (see Figure 2 above).

The CNIL has fined companies using AI systems in a number of domains, namely: data security, personalised advertising, facial recognition, and data brokerage. Table 1 below offers an overview of the main decisions issued by the CNIL related to activities involving the use of AI since the GDPR came into force. Although the overview is non-exhaustive, it contains the most important decisions based on the fines involved. It is also representative of the focus given by the CNIL in its enforcement action, for instance with regard to personalised advertising.

Table 1: CNIL decisions sanctioning entities intensively using AI systems (2018–2024)

Date	Domain	Entity	Fine (in EUR)
04/04/2024	Data brokerage	Hubside.store	525,000
31/01/2024	Data brokerage	Foriou	310,000
29/12/2023	Data brokerage	Tagadamedia	75,000
29/12/2023	Personalised advertising	Yahoo	10 million
29/12/2023	Personalised advertising	NS Cards	105,000
15/06/2023	Personalised advertising	Criteo	40 million
08/06/2023	Personalised advertising	KG COM	150,000
11/05/2023	Personalised advertising	Doctissimo	380,000
17/04/2023	Facial recognition	Clearview AI	5.2 million
29/12/2022	Personalised advertising	Voodoo	3 million
29/12/2022	Personalised advertising	TikTok	5 million
29/12/2022	Personalised advertising	Apple	8 million
19/12/2022	Personalised advertising	Microsoft	60 million
17/10/2022	Facial recognition	Clearview AI	20 million
31/12/2021	Personalised advertising	Facebook	60 million
31/12/2021	Personalised advertising	Google	150 million
27/07/2021	Personalised advertising	Le Figaro	50,000
07/12/2020	Personalised advertising	Amazon	35 million
07/12/2020	Personalised advertising	Google	100 million
18/11/2020	Personalised advertising	Carrefour	3.05 million
21/01/2019	Personalised advertising	Google	50 million
19/12/2018	Data security	Uber	400,000
24/07/2018	Data security	Dailymotion	50,000

Source: CNIL

As this overview shows, the CNIL has indeed given priority to enforcing the GDPR with respect to personalised advertising, which relies heavily on AI systems to operate. For several years, the CNIL has prioritised investigations related to cookies and dark patterns, i.e., deceptive designs of user interfaces leading users to accept things they would not otherwise have agreed to, such as the placing of cookies on their devices for personalised advertising purposes.

In its administrative practice, the authority has relied on the latest research, also based on findings in computer science. In 2020, an empirical study showed that 'cookie banners' appearing on websites and seeking consumer consent to place cookies on their devices were generally not compliant with the GDPR: just 11.8% of websites using the five top consent management platforms met minimal GDPR requirements for valid consent (Nouwens et al., 2020).

The French data protection authority inter alia referred to this empirical study in its decision to fine Facebook EUR 60 million for the use of dark patterns, namely not enabling users to reject cookies as easily as accepting them (decision SAN-2021-024). Discouraging users to decline cookies while encouraging them to accept being tracked on the first page undermined their freedom of consent as many users would not accept cookies if offered a genuine choice.

The CNIL also fined Google EUR 150 million for similar practices (decision SAN-2021-023). Applying the French law transposing the ePrivacy Directive (2002/58/EC) in light of the heightened consent requirements under the GDPR, the authority held that the method employed by Google Search and YouTube for users to manifest their choice over the placing of cookies was illegally biased in favour of consent. Again, the authority referred to several studies showing that organisations implementing a 'reject all' button on the first-level consent interface had seen a drop in the consent rate to accept cookies. The CNIL also relied on the same studies to impose heavy fines on Microsoft (decision SAN-2022-023) and TikTok (decision SAN-2022-027).

The CNIL focused not only on deceptive designs that manipulate users into sharing more personal information than they intended but also on other types of invalid consent or the lack thereof.

For instance, in 2019 the CNIL imposed a EUR 50 million fine on Google for failing to obtain valid consent (i.e., informed, unambiguous, specific) from its users for purposes of personalised advertising (decision SAN-2019-001), later upheld by the Conseil d'Etat (decision n°430810). The regulator determined that user consent was not sufficiently informed because the information about the company's processing of personal data for ad personalisation was not centralised in a single document. As a result, users were unaware that their information would be processed for this purpose across Google's various services. Nor was consent unambiguous given that the option to be shown personalised ads was pre-checked in difficult-to-find settings. Finally, nor was consent specific, because Google sought consent when users were creating an account, by ticking boxes to agree on its terms of service and privacy policy, whereas consent should be obtained distinctively for each purpose. Since then, Google adapted its practices across the EU to comply with GDPR requirements.

More recently, the CNIL fined Yahoo EUR 10 million for placing advertising cookies on users' devices without collecting their prior consent at all (decision SAN-2023-024). The CNIL found that requiring non-essential cookies for service use is not illegal if users can freely give or withdraw consent without negative consequences. However, the authority noted that withdrawal was difficult due to service interruptions, a lack of alternatives, and misleading interface elements that complicated the process for users. Similarly, the authority fined newspaper Le Figaro EUR 50,000 for placing advertising cookies on its website without obtaining user consent (decision SAN-2021-013).

Overall, just between 2020 and 2021, the CNIL adopted around 70 corrective measures (formal notices and sanctions) related to non-compliance with cookie regulation, especially related to consent (CNIL, 2021). In total, 80% of the affected entities rapidly complied with the authority's orders, with effects not only in France but at least across the EU.

If advertising might not be the first thing that pops to mind while thinking about AI, it has played a key role in how the digital economy has developed, with relevance in today's developments around generative AI. Advertising has been described as the "lifeblood of the internet" (Cofone & Robertson, 2018: 1472) because it has enabled business models offering free services, making possible the development of some of the largest online platforms, such as those operated by Google and Meta. Many of those platforms have now been designated as gatekeepers under the DMA due to their important market power and ability to raise barriers to entry for potential competition. Most fines imposed by the CNIL in this space in fact relate to the illegal placing of advertising cookies linked to both companies. In turn, companies like Alphabet, Amazon, Microsoft and Meta – which have relied on scores of personal data and engaged in personalised advertising – are now among the main developers of, and investors in generative AI models (Autorité de la concurrence, 2024). Therefore, the link between personalised advertising and further AI development is more important than usually believed, yet it lies at the core of the activities of some of the most valuable companies on the planet.

Finally, the CNIL's early GDPR enforcement has also focused on another domain relevant for its interaction with the AI Act: facial recognition. In 2022, Clearview AI, which had amassed over 20 billion images for its facial recognition service by scraping publicly accessible websites, was fined EUR 20 million by the CNIL for multiple GDPR violations (decision SAN-2022-019). The investigation revealed Clearview's unlawful data processing without user consent, and failure to respect individual data rights of access and erasure. Clearview AI also failed to cooperate with the CNIL in this initial investigation, and after it was ordered to cease data collection in France and delete existing data. The company thus faced additional daily penalties for non-compliance, totalling EUR 5.2 million in 2023 (decision SAN-2023-005).

At any rate, the AI Act will now explicitly prohibit AI systems that create or expand facial recognition databases via the untargeted scraping of facial images from the Internet or CCTV footage, such as those offered by Clearview AI.

4.2 Application of Data Protection Rules to Generative AI Systems

Since the release of advanced generative AI applications in late 2022, the CNIL has started working on applying the GDPR to these new systems. So far, it has largely focused on policy work, as no sanctions have yet been imposed on generative AI companies.

In 2023, the authority unveiled an AI action plan, extending its policy and regulatory work to generative AI and large language models (CNIL, 2023a). In its plan, the CNIL recognises the importance of personal data in this space, focusing on transparency, data security, bias prevention, and the ethical use of AI technologies. Building on years of prior work in the AI field, the CNIL's agenda expands to include generative AI, chatbots, and other derivative applications. The plan is structured around four key objectives: (i) understanding AI systems and their impacts; (ii) ensuring privacy-respecting AI development; (iii) fostering innovation within the AI ecosystem in France and Europe; and (iv) and auditing and regulating AI systems to protect individuals. The plan includes publishing guidelines (many of which have been subject to public consultation), engaging with AI developers, and conducting audits. Overall, its action plan reflects a comprehensive approach to balance innovation with privacy and ethical considerations.

However, France is not the only member state where data protection authorities have initiated discussions and enforcement actions regarding generative AI applications under the GDPR.

Important concerns relate to the very development of generative AI models. In May 2024, the Dutch data protection authority (AP) issued guidelines related to web scraping, the automatic collection and storage of online information (Autoriteit Persoonsgegevens, 2024). The authority determined that this practice is almost always illegal due to privacy risks and GDPR violations. This is especially relevant for generative AI applications that rely on large datasets as they often involve the collection of personal data without consent. The AP emphasised that publicly accessible information does not imply permission for scraping, and exceptions are rare, typically limited to non-commercial, personal projects or highly targeted corporate uses. These findings raise doubts as to the compatibility of web scraping involving personal data with the GDPR, albeit the issue is not settled yet.

In fact, the Italian data protection authority is still conducting investigations against OpenAI for alleged GDPR breaches, following a temporary ban of its ChatGPT chatbot in March 2023. The suspected violations include illicit personal data collection via web scraping without user consent or a proper legal basis.

Still, privacy concerns have also arisen regarding the training of AI models based on user-generated content on social media. Meta's announcement in May 2024 that it was updating its privacy policy to allow for the training of its AI models with user data sparked controversy over GDPR compliance. In its updated policy, Meta claimed that it had legitimate interests to train its AI models on the content users generated on Facebook and Instagram, including personal data, thus justifying bypassing user consent for this type of data processing. However, following several GDPR complaints – including before the CNIL, Meta postponed its AI features in Europe (Meta, 2024). Putting pressure on regulators, Meta cited concerns with innovation and competitiveness on the continent.

Further privacy issues relate to the use of generative AI systems when they have already been deployed. In its investigations into OpenAI, the Italian data protection authority is concerned that ChatGPT's lack of an age verification system exposes minors to inappropriate content. Other concerns relate to ChatGPT's tendency to produce false information ('hallucinating'), particularly as regards individuals, holding significant implications for the GDPR's obligations on personal data accuracy and user rights in this respect. In April 2024, the Austrian data protection authority received a complaint directed at OpenAI for the inherent inaccuracies and lack of transparency in data generated by ChatGPT. Despite requests for data access and rectification, OpenAI contends that it cannot rectify generated data, thus potentially infringing upon GDPR provisions.

These recent developments in regulatory scrutiny over generative AI show the need for a common European approach to AI regulation. The CNIL and other data protection authorities have indeed started cooperating in this space, as the next section explains.

5 Cooperation and Collaboration in AI Regulation

Collaboration in shaping privacy-friendly AI regulations has been taking place in both the adoption of legislation, notably the AI Act, and in enforcement actions.

In France, multiple institutions have contributed to shaping AI regulation in a way that considers data protection issues, while also collaborating with their European counterparts. The CNIL was cooperating with a number of stakeholders on AI regulation even before the uptake of generative AI applications. In 2018, the CNIL already stated that it was actively involved in shaping international guidelines and ethical standards for AI development (CNIL, 2019: 31). A few years later, it was involved in shaping the AI Act based on the European Commission's proposal of 2021. Indeed, the authority collaborated with its European peers on the European Data Protection Board (EDPB) to assess the proposal and make recommendations. This cooperation resulted in the publication of an opinion in which data protection authorities stressed the important overlaps between AI and data protection regulation and the challenges in aligning the AI Act with the GDPR (EDPB-EDPS, 2021).

In 2022, the Conseil d'Etat also published a study on AI governance under the AI Act in which it recommended that the CNIL become the national supervisory authority under the Act, implying its profound transformation and increased resources (Conseil d'Etat, 2022). In line with these recommendations, the CNIL established a new department focusing on AI in early 2023. Initially comprising five experts in law and engineering, this new department aims to enhance the CNIL's understanding of AI systems and address privacy risks. Apart from supporting other departments, it provides guidance on legal compliance, and collaborates with the EDPB ahead of the AI Act's application.

Collaboration also took place in enforcement action, especially regarding generative AI. Given the many privacy issues arising from the development and deployment of generative AI systems, various data protection authorities within the EDPB have established a taskforce dedicated to interpreting the GDPR rules applicable to ChatGPT. Its mandate included the exchange of information, the coordination of external communication by different data protection authorities in their enforcement activities, and the identification of issues for which a common approach is needed in the context of their different enforcement actions concerning ChatGPT. This taskforce published a report in May 2024, which also mentioned the great importance of providing further guidance on the interactions between the GDPR and the AI Act (EDPB, 2024). However, because the investigations mentioned above are still ongoing, the report only contains preliminary views on certain aspects of the cases.

6 Regulatory Sandboxes and Innovation

In line with its AI action plan, the CNIL seeks to support innovation in the French and European AI ecosystem while upholding fundamental rights, especially privacy. Key initiatives include a 'sandbox' programme, initially guiding projects in health and education (in 2021 and 2022, respectively), some of which relied on AI. Here, the focus has mainly been to ensure innovative projects comply with GDPR requirements. Apart from its sandbox initiatives, the CNIL provided support for augmented video surveillance providers for the 2024 Olympics and launched a new programme assisting companies with GDPR compliance. The CNIL also aims to maintain ongoing dialogue with research teams, R&D centres, and AI companies to ensure data protection compliance (CNIL, 2023a).

While the GDPR did not offer a specific framework for regulatory sandboxes – although it encourages the development of codes of conduct and certification, the AI Act now establishes one for innovative AI projects. The aim of these sandboxes is to provide a safe testing environment for AI providers to experiment with AI systems before they are put on the market. Ultimately, the objective is that regulatory supervision ensures adherence to standards like transparency, accuracy and non-discrimination, as well as legal certainty for innovators. Within this framework, cooperation between member states is encouraged, constituting further cooperation opportunities for data protection authorities like the CNIL.

Of course, the development of AI systems most generally involves complex data protection issues. Therefore, the AI Act mandates that national authorities involve data protection authorities in overseeing AI sandboxes, especially when personal data is processed. In fact, the CNIL had already established AI sandboxes ahead of the Act's application. In 2023, it opened a new call for AI projects focused on public services in which participants benefit from the expertise of its AI department (CNIL, 2023b). Eight projects have been selected, spanning public administration, employment services, environmental services, public transport, healthcare, legal and administrative information, and sports performance analysis. It remains to be seen whether the CNIL will have sufficient resources once the AI Act applies to further expand its sandbox programmes, especially given the regulation's aim that "AI regulatory sandboxes should be widely available throughout the Union".

Finally, the AI Act allows AI providers to use personal data lawfully collected for another purpose to develop, train or test certain AI systems in those sandboxes, but only if they serve public interests (e.g., public safety, public health, environment, energy sustainability, transport, or public administration). In order to avoid inconsistencies with the GDPR (EDPB-EDPS, 2021), for example the principle of purpose limitation, the AI Act outlines strict conditions in which this personal data can be used within regulatory sandboxes. For instance, it foresees safeguards for sensitive data and explicitly provides that this possibility should not constitute an exemption to the right not to be subject to a decision based solely on automated processing, including profiling.

7 Conclusion

This contribution argues that, beyond the recent hype surrounding generative AI and the adoption of the AI Act, AI systems have long been subject to important legal limitations. This has been the case in France and across Europe.

Given the centrality of (personal) data for the development and use of AI systems, data protection regulation has played a key role in regulating them, notably through the Data Protection Directive and the GDPR. As discussed here, national data protection authorities such as France's CNIL have been playing a major role in regulating algorithmic practices. Under the AI Act, the role of data protection authorities will be expanded. As explained, the CNIL was already preparing for this transition, but more resources are probably needed for effective enforcement. In addition, competition and copyright law have also, and will continue to substantially impact the development and use of AI systems, including those based on foundation models.

Yet, regulatory gaps remain and are sometimes filled by court rulings. At the EU level, this was true regarding the 'right to be forgotten' which the CJEU initially established. In France, the Conseil constitutionnel has even applied the Declaration of the Rights of Man and of the Citizen of 1789 to fill such a gap. Establishing a constitutional right to access administrative documents, the court specified how universities must reveal the criteria and potential algorithmic methods used to evaluate student applications (decision 2020-834 QPC). Despite the enthusiasm with the AI Act, legislation is therefore not a panacea, but must sometimes be complemented by constitutional rules and principles.

Finally, regulating technology – especially AI systems – requires continuous updates and adaptability in laws. The AI Act foresees possibilities of adaptation already. But it is key for regulators to keep cooperating (within Europe and globally), for instance on standards, in order to foster innovation while guaranteeing a high level of protection for consumers and citizens.

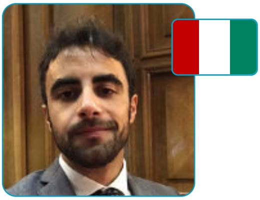
REFERENCES

- Autoriteit Persoonsgegevens (2024). AP: scraping bijna altijd illegaal. Retrieved 27 June 2024, from https://autoriteitpersoonsgegevens.nl/actueel/ap-scraping-bijna-altijd-illegaal.
- Autorité de la concurrence (2024). Avis 24-A-05 du 28 juin 2024 relatif au fonctionnement concurrentiel du secteur de l'intelligence artificielle générative. Retrieved 2 July 2024, from https://www.autoritedelaconcurrence.fr/sites/default/files/integral_texts/2024-06/avisIA.pdf.
- Bradford, A. (2021). The Brussels effect: How the European union rules the world. Oxford University Press.
- CNIL (2024). Cahier air2023. IA et libre-arbitre: sommes-nous des moutons numériques? Retrieved 27 June 2024, from https://www.cnil.fr/sites/cnil/files/2024-04/cahier_air2023.pdf.
- CNIL (2023a). Intelligence artificielle : le plan d'action de la CNIL. Retrieved 27 June 2024, from https://www.cnil.fr/fr/intelligence-artificielle-le-plan-daction-de-la-cnil.
- CNIL (2023b). « Bac à sable » intelligence artificielle et services publics : la CNIL accompagne 8 projets innovants. Retrieved 27 June 2024, from https://www.cnil.fr/fr/bac-sable-intelligence-artificielle-et-services-publics-la-cnil-accompagne-8-projets-innovants.
- CNIL (2021). Refuser les cookies doit être aussi simple qu'accepter : bilan de la deuxième campagne de mises en demeure et actions à venir. Retrieved 27 June 2024, from https://www.cnil.fr/fr/refuser-les-cookies-doit-etre-aussi-simple-quaccepter-bilan-de-la-deuxieme-campagne-de-mises-en.
- CNIL (2019). Rapport d'activité 2018. Retrieved 27 June 2024, from https://www.cnil.fr/sites/cnil/files/atoms/files/cnil-39e_rapport_annuel_2018.pdf.
- CNIL (2017). Concertation citoyenne sur les enjeux éthiques liés à la place des algorithmes dans notre vie quotidienne : synthèse de la journée. Retrieved 27 June 2024, from https://www.cnil.fr/sites/cnil/files/atoms/files/cr_concertation_citoyenne_algorithmes.pdf.
- Cofone, I. & Robertson, A. (2018): Consumer Privacy in a Behavioral World. Hastings Law Journal, 69(6), 1471-1508. https://repository.uchastings.edu/hastings_law_journal/vol69/iss6/1.
- Conseil d'Etat (2022). S'engager dans l'intelligence artificielle pour un meilleur service public. Retrieved 27 June 2024, from https://www.conseil-etat.fr/actualites/s-engager-dans-l-intelligence-artificielle-pour-un-meilleur-service-public.
- EDPB (2024). Report of the work undertaken by the ChatGPT Taskforce. Retrieved 27 June 2024, from https://www.edpb.europa.eu/system/files/2024-05/edpb_20240523_report_chatgpt_taskforce_en.pdf.
- EDPB-EDPS (2021). Joint Opinion 5/2021 on the proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Retrieved 27 June 2024, from https://www.edpb.europa.eu/system/files/2021-06/edpb-edps_joint_opinion_ai_regulation_en.pdf.

- EU Agency for Fundamental Rights (2024). Lack of resources undermine EU data protection enforcement. Retrieved 27 June 2024. https://fra.europa.eu/en/news/2024/lack-resources-undermine-eu-data-protection-enforcement.
- European Commission (2024). Speech by EVP Margrethe Vestager at the European Commission workshop on “Competition in Virtual Worlds and Generative AI”. Retrieved 2 July 2024, from https://ec.europa.eu/commission/presscorner/detail/en/speech_24_3550.
- European Commission (2021). 2030 Digital Compass: the European way for the Digital Decade. COM(2021) 118 final.
- European Commission (2020a). White Paper on Artificial Intelligence - A European approach to excellence and trust. COM(2020) 65 final.
- European Commission (2020b). Shaping Europe’s Digital Future. Retrieved 27 June 2024, from https://commission.europa.eu/system/files/2020-02/communication-shaping-europes-digital-future-feb2020_en_4.pdf.
- European Commission, Legal service (2022). 70 years of EU law : a Union for its citizens, Publications Office of the European Union. https://data.europa.eu/doi/10.2880/02622.
- Hacker, P., Cordes, J., & Rochon, J. (2024). Regulating Gatekeeper Artificial Intelligence and Data: Transparency, Access and Fairness under the Digital Markets Act, the General Data Protection Regulation and Beyond. European Journal of Risk Regulation, 15(1), 49-86. https://doi.org/10.1017/err.2023.81.
- Margoni, T., & Kretschmer, M. (2022). A deeper look into the EU text and data mining exceptions: Harmonisation, data ownership, and the future of technology. GRUR International, 71(8), 685-701. https://doi.org/10.1093/grurint/ikac054.
- Meta (2024). Building AI Technology for Europeans in a Transparent and Responsible Way. Retrieved 27 June 2024, from https://about.fb.com/news/2024/06/building-ai-technology-for-europeans-in-a-transparent-and-responsible-way/.
- Milmo, D. (2023, February 2). ChatGPT reaches 100 million users two months after launch. The Guardian. https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app.
- Nouwens, M., Liccardi, I., Veale, M., Karger, D., & Kagal, L. (2020). Dark patterns after the GDPR: Scraping consent pop-ups and demonstrating their influence. Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3313831.3376321.
- OECD (2024), AI, data governance and privacy: Synergies and areas of international co-operation. OECD Artificial Intelligence Papers, No. 22. https://doi.org/10.1787/2476b1a4-en.

Chapter 3

The Italian Draft Law on Artificial Intelligence Systems: Key Differences and Complementarities with the EU AI Act



Gian Marco Bovenzi

1 Introduction

At a press conference held on 23 April 2024, Italian Minister of Justice Carlo Nordio presented the new "Draft Law on Artificial Intelligence Systems". The draft law has not become a law yet since not all the steps of the legislative iter have been completed and it is still pending before the Parliament. However, if approved, the bill would become the first national law among EU member states to exclusively regulate AI systems – and, while there might be minor amendments in the final version, the chances of it being approved are very high, also given the clear governmental majority in the Parliament.

Nowadays, the regulation of artificial intelligence (AI) is viewed as a priority. The use and role of AI not only within industry but in everyday life is become increasingly central, and specific regulations are called for in order to prevent and avoid possible misbehaviours associated with its implementation.

Of course, the European Union is the frontline. Not only the recently enforced Artificial Intelligence Act (hereinafter "AI Act") passed in the European Parliament on 13 March 2024 and was formally adopted by the Council of the European Union on the followig 21 May 2024, but also several other regulations in the field underlie the pivotal interest of EU institutions in regulating generative AI: let us consider Regulation (EU) 2022/2065 (Digital Services Act – DSA), Regulation (EU) 2022/1925 (Digital Markets Act – DMA), Regulation (EU) 2022/868 (Digital Governance Act – DGA), but also the Data Act of 13 December 2023, n. 2854/2023, and other proposed directives and regulations in the field.

Nevertheless, an exclusively EU-based AI regulatory framework is not enough. Although the EU Regulations are directly applicable in member states' legislative frameworks, and thus are self-executive directives, this might not be sufficient to effectively regulate the effects of the use of AI in the European territory. States need to develop their own frameworks, which of course should not overlap and clash with AI Act, in order to harmonise the various items of legislation across the EU.

In this respect, as we have seen in the first part of this introduction, Italy has actually taken steps. As the next sections highlight, the draft law does not cover every field of artificial intelligence, and the regulatory gaps are perhaps far from being fully filled: nonetheless, advancement of the bill implies that the government includes the regulation of AI among its priorities. In this paper, a general assessment of the draft law is first given, stressing the main areas of its application and contents (paragraph 2 and following sub-paragraphs); in paragraph 3, a comparative analysis of the draft law and the AI Act is undertaken; finally, the conclusions point to existing gaps (either between the two regulations or generally speaking) and provide policy recommendations for filling such gaps and excluding possible threats due to their non-regulation.

2 The Italian Draft Law on Artificial Intelligence Systems

The Italian draft law on artificial intelligence systems contains 26 articles and its clear goal is to promote the full exploitation of the potential of AI. To achieve these objectives, it is pivotal to ensure the more responsible use of AI without compromising security and individual rights. The Council of the Government itself (in Italian "Consiglio dei Ministri", basically representing the government under the Constitution) declared that *"The draft law identifies regulatory criteria capable of balancing the relationship between the opportunities offered by the new technologies and risks tied to their implementation, to their under-usage and their harmful use. Moreover, it introduces those principles that one on hand promote the use of new technologies to enhance the citizens' life conditions and social cohesion and, on the other hand, provide solutions for risk management based on an anthropocentric vision"* (Consiglio dei Ministri, 2024).

The government has also explicitly declared that the draft Law does not overlap with the European Union AI Act, but *"goes along with its regulatory framework, filling the spaces reserved to national legislation, considering that the Regulation is based on the architecture of the risks related to the used of Artificial Intelligence".*

Summing up, the Italian government's clear objective is not to replace, but to harmonise the Italian AI regulatory framework with that of the European Union (Senato della Repubblica, 2024). In order to understand whether the goal can actually be said to be 'achieved', below we analyse the proposed legislation.

The draft law includes specific regulations on several sectors, as analysed in the following sub-paragraphs. Before analysing these sectors, it is useful to briefly recall the draft law's general principles since it aims at:

- Promoting the use of new technologies based on the respect of fundamental rights, principles and liberties, and the rule of law;
- Fostering the social cohesion and accessibility of AI systems;
- Providing cyber-secure risk-management solutions based on an anthropocentric and non-discriminatory vision; and
- Regulating the use of AI in the labour market and in the judiciary/courts.

In order to achieve the mentioned goals, the draft law underlies the priority given to several principles (Article 3), which may be synthesised as:

- The interest in fair and correct algorithms: here, the research, experimentation, development, adoption and application of algorithms must be implemented with respect to citizens' fundamental rights and liberties, and respecting the further principles of transparency, proportionality, security, data protection, accuracy, non-discrimination, sustainability, and inclusivity;
- The interest in data protection: here, the development of AI systems and models must abide by the principles of the proportionality, fairness, reliability, security, quality, and transparency of the data – here, cybersecurity plays a fundamental role; and
- The interest in digital sustainability: here, the development and concrete application of AI systems and models must be deployed in respect of a person's autonomous decision and power, preventing potential harm, and without prejudice to the integrity of the State and its institutions.

Moreover, the draft law:

- Provides regulations applying to economic development and investments in the sectors of industry and start-ups, thereby expecting to boost the AI economy and market in the country and making Italy a leader in the sector – all while respecting the EU principles and framework on fair market competition;
- Stresses the pivotal role of information and data protection: all AI systems must respect the existing legal framework for privacy and data protection, also with reference to telecommunications systems;
- In terms of governance, it sets up two national authorities for AI, the Agency for a Digital Italy ("AgID") and the Agency for National Cybersecurity ("ACN"), with the goal of ensuring the creation and joint management of new experimental models finalised upon realising AI systems in compliance with both national and EU regulations, as well as assuring efficient multilevel governance; and
- Constantly recalls the need to align the national system to the EU AI Act in citizens' literacy and professional training, but also in terms of high-risk AI systems' accountability and transparency: in short, the Italian draft law on AI and the EU AI Act do not overlap, but are instead complementary and harmonised, fostering multilevel cooperation.

In conclusion, the bill appears to cover several areas not only of the law, but of state governance since it includes economic measures, incentives for investments, data protection, and the foundation of new national authorities specially dedicated to the surveillance of the correct deployment of AI systems and models and their compliance with the existing regulations. Recognising the supremacy of the EU AI Act, all of this is therefore intended to harmonise the two regulations, without causing any overlap.

In order to fully assess the draft law – and be able to draw a critical line through it – we now analyse the sectors expressly covered by the proposed framework.

2.1 Health and Disability

As general principles, the draft law establishes that in those cases in which AI will be used in health and medical treatments, hospitals and medical structures are obliged to inform the citizens of the technology used in the treatment. Further, research and experimentation of AI systems in the health domain is considered to be in the public interest and, finally, AI will support cure and assistance across the whole territory.

More specifically, this section highlights the importance of the role of AI in the health sector. In fact, AI carries the possibility of bringing revolutionary benefits in this field, making treatments more adequate and 'smarter' also via the use of systems that are better tailored to the needs and urgency of each patient. The goal is thus to enhance the whole health system and prevent illnesses: all this of, respecting the principles of non-discrimination, accessibility and inclusiveness, also when it comes to patients with disabilities and access to the national health plans and services, should be enhanced with the use of AI systems.

It is interesting to deepen the provision according to which research and experimentation related to AI systems is considered to be in the public interest: this implies that the use of personal data in this field be considered and protected as such (also under the GDPR) and, hence, the obligation for the hospital or medical structure to ask for the patient's agreement is not required, whether the use of personal data was initially authorised for another research – of course, also related to the medical field.

Overall, the section of the draft law dedicated to health and disability does not change the current framework as much – since it is limited to defining AI systems (and their use) in the field, as well as the public interest entailed in AI-related research. Still, it represents an effort to clarify goals and objectives in the field.

2.2 Labour Law

The rationale of the use of AI in labour law and within the job market is to enhance working conditions, offering solutions based on risk management, and respecting the principles of fairness and non-discrimination, as well as to exploit the social value of digital technologies.

Given the peculiar role of human dignity within the labour law, the draft law underlines that use of AI in working places must respect such principles, the worker has to be informed about the use of an AI system in the workplace and, of course, such use must respect the principles of privacy and data protection. In order to evaluate the fair and non-discriminatory use of AI in working places, the draft law establishes an Observatory on the adoption of AI systems in workplaces. Finally – and this provision appears relevant – in the field of intellectual professions (e.g., lawyer) the use of AI must be limited, and prevalence is accorded to human critical thought.

Aside from the provisions concerning privacy and respect for the workers' dignity, it is interesting to stress the provisions about the prevailing role of human critical thought over AI in intellectual professions. In fact, several instances and potential issues have recently been raised over the role of generative AI in several working domains, perhaps mostly in the legal professions where, possibly, legal acts might be generated by AI systems such as ChatGPT. Therefore, the provision stating that critical thinking must always be prevalent appears to be a clear statement in tackling the use of AI systems in certain jobs.

2.3 Public Administration

Within the public administration, the role of AI should be to guarantee the administrative and bureaucratic constitutional principles of efficiency and impartiality. Yet, at the same time, AI should respect the principles of self-determination and human responsibility, with this implying that, even though a civil servant uses AI systems in tehri administrative activities, the liability for its actions should never be given to AI systems.

Although the provisions on public administration are quite scarce compared to other fields, it is nevertheless worth highlighting that stressing the principle of personal responsibility for actions taken using AI systems implies that the rationale is not to let AI systems 'overcome' individual activities.

2.4 Judiciary

judiciary perhaps represents the most sensitive sector when it comes to the use of AI systems. Certainly, AI can represent an important and efficient tool for assisting the work of judges, magistrates, prosecutors and clerks, but it should never be forgotten that the interpretation of the law, the evaluation of proof and evidence in a trial, as well as the motivated decision of a sentence are activities than can be exclusively and undoubtedly performed with the reasoning of a human being's. AI systems are incapable of deductive reasoning, and when considering the thought and instances raised about the possibility of AI systems replacing the judiciary, the Minister himself declared that "AI systems are used exclusively for the organisation and the simplification of the judicial workload and the research in caselaw, jurisprudence, or doctrine [...] a judge shall not be conditioned [by an AI system]".

Let us not forget that the use of AI in the judiciary is considered to be high-risk under the AI Act given its potentially relevant impact on democracy, the rule of law, individual liberties, and the right to a fair trial: as such, provisions referring to high-risk systems apply. Moreover, the European Commission for the Efficiency of Justice (CEPEJ) recently adopted the European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment.

The European legal framework is thus strongly committed to ensuring that AI systems do not interfere with the sphere of the judiciary. The Italian draft law is also consistent with such principles: the proposed law is clear in defining the boundaries between the roles of AI and of judicial operators, leaving AI only with the role of supporting the activities of judiciaries in simplifying workloads and conducting research in case law, jurisprudence and doctrine, leaving the magistrate with the exclusive role of interpreting the law, evaluating proof and evidence in a trial, as well as to provide motivation to their judgments – thereby aligning with European and international frameworks in the field.

2.5 National Cybersecurity

National (cyber)security is a pivotal objective in the regulation of AI systems and tools. The most relevant novelty introduced with the draft law is the introduction of a National Strategy for AI, whose results have to be communicated to the Parliament year after year. Further, as mentioned in paragraph 2, the draft law sets up two National Authorities for AI, the Agency for a Digital Italy ("AgID") and the Agency for National Cybersecurity ("ACN"), whose role is fundamental for ensuring compliance with the regulations and introducing new opportunities for innovation in the field of technology.

Overall, the provisions of the draft law are necessarily to be harmonised with other existing regulations and legal frameworks currently enforced in Italy. The draft law does not overlap, but integrates such provisions on a multi-level perspective.

2.6 Criminal Law

In the field of criminal law, the draft law introduces different measures:

- Harsher sentences for several crimes committed with the use of AI systems, considered as an aggravating circumstance in light of the insidious nature of an AI system – given the huge potential consequences of the use of generative AI. The crimes to which the circumstance applies are expressly listed in the draft law (among others, computer fraud and money laundering);
- A new crime, that is, the punishment of those who release and broadcast pictures, videos or language that are falsified and created using AI systems, causing unfair damage to a person; and
- 'Aggravating circumstances with a special effect' for several crimes, meaning (again) that the punishment would be harsher when criminal conduct is committed with the use of AI systems.

Moreover, the draft law includes an explicit recommendation for the government to adopt following its adoption, referring to several further laws that aim at criminalising illicit conduct committed with the use of artificial intelligence. Such conduct still has to be identified, depending on the level of harm potentially caused by an AI system to a person. This is due to the fact that AI holds the capacity to generate contents – and thus, to foster conduct – that seems real, but is actually virtual and, as such, considered to be insidious and harmful.

Overall, by including new crimes or aggravating circumstances in the existing Italian criminal code the draft law acknowledges the possible harm arising from unregulated or negligent use of AI systems. Criminal law is a peculiar branch of the law: as we know, the European Union does not have competence with respect to regulating member states' criminal systems because the criminal exigences vary country by country. Further, the (huge) number of fundamental principles underlying criminal systems make it extremely demanding to legislate in this field, also since an AI system could basically 'shift' the conduct from a person to a machine – thereby potentially infringing on the principle of criminal intent in criminal conduct that underpins not only states' criminal frameworks, but also the basic principles of the EU when it comes to fundamental rights.

In conclusion, the draft law tries to regulate several criminal aspects of the use of AI systems and punishes those who use them insidiously and cause harm to other persons; nevertheless, many steps still have to be taken, and what the proposed law regulates so far cannot be regarded as exhaustive.

> ”
> AI holds the capacity to generate content that seems real but is actually virtual.

2.7 Intellectual Property and Copyright Law

According to several scholars and researchers, intellectual property and copyright law are among the branches most 'at risk' of the law when it comes to the use of AI. This viewpoint is expressed in the context of the peculiar nature of generative AI, which is capable of simulating and replicating voice, image, text and audiovisual material generally. As we know, ChatGPT can create a whole book, a short novel, or carry out complete research. All of this is capable of infringing upon IP and copyright law if left unregulated.

The draft law aims at *"fostering the identification and recognition of the use of AI systems in the generation of textual, photographic, audiovisual, and radiophonic contents"*. However, at the same time it also seeks to protect the copyright for works created with AI systems. The draft law therefore adopts a two-side approach: on one hand, AI-generated works must be recognised and possibly turned down when they infringe on others' works; on the other hand, if a work is new, it has to be protected even if made with the use of AI due to the increasing production of works made with the help of generative AI systems. It goes without saying that such works have to have been previously unreleased and not covered by other authors' copyrights and intellectual property.

3 Congruities and Differences Between the Italian Draft Law and the EU AI Act

We have thus far provided an in-depth assessment of the Italian draft law on AI systems, highlighting how the proposal is the first effort not only of Italy, but of any member state, to set out the exhaustive regulation of artificial intelligence within different sectors and branches of the law (Sorrentino, 2024) – including peculiar and sensitive ones like criminal law.

Nevertheless, member states' legislative framework – necessarily – has to be harmonised with the EU regulatory framework, and cannot clash with its provisions (unless in cases where competences are exclusively left to member states). In fact, in Article 22 the draft law explicitly mentions the need for the government to adopt further laws, law decrees, legislative decrees, or other acts with the force of law, harmonising Italy's framework with the EU AI Act. Such provision underlines the need to ensure the two legal frameworks do not overlap – which would basically result in the non-applicability of the Italian law given the hierarchical priority of EU law – but are complementary. This will possibly be achieved in the next few years with several governmental measures to be adopted, such as:

- Designating Italian local authorities to be competent for supervising the complementarity of the two regulations, and to be a constant point of contact between Italy and EU institutions, and empowered to enforce the provisions of the AI Act in the country;
- Adopting measures aiming at fostering digital and AI literacy among the population and Italian institutions alike;
- Applying such measures seeking to foster digital literacy also among professionals; and
- Including digital and AI literacy in schools and universities, specifically within the STEM sectors (Science, Technology, Engineering, Mathematics) so that students – and future workforce – are aware of the effects of technology in society and can use such technology in their work.

Complementarity between the Italian draft law and the AI Act is possible.

When we (briefly) analyse the AI Act's provisions, we notice how they do not cover, at least not exhaustively, the sectors regulated by the Italian draft law. The AI Act has the objective of creating a harmonised regulatory framework in the field of artificial intelligence that is applies to all members and, in so doing, it (generally) stresses the need for artificial intelligence systems to be developed and used safely, ethically, democratically, without infringing upon citizens' fundamental rights.

The biggest risk associated with AI systems implied by the AI Act is that their non-regulated use could pose significant threats and harm not only to citizens, but to society, economic systems, security, and other crucial sectors for states. AI is thus clearly a pivotal instrument for fostering and promoting innovation, industrial competitiveness, and also for bringing benefits for citizens' well-being while, at the same time, consumer, worker and citizen protection in general is required. In order to achieve such protection, service providers and companies must respect several obligations depending on the apparent level of risk that an AI system can bring: the higher the risk, the greater the obligations.

In general, the main objectives of the AI Act are to:

- Create a common EU market for AI, encouraging the free circulation of AI systems in EU territory provided that they are safe and abide with the principles of ethics, security, democracy, and respect of citizens' fundamental rights as mentioned above;
- Enhance citizens' trust in AI systems, ensuring that they are their affordable, trustworthy and transparent, in order not to pose harm and risks to the

citizens when it comes to their use. This is a critical aspect: much has already been said about the possibility, in a (distant?) future, of AI and robots overcoming humans. Although this is a dystopian scenario, it is likewise true that such narrative has for many citizens led to ideas that the empowerment of AI might bring catastrophic consequences for society. In turn, assuring that AI systems are trustworthy and reliable represents a crucial objective so that people use AI safely, without being 'scared' of it, and instead learn how to deal and live with it since it is undeniable that AI is at once the present and the future;

- Prevent and mitigate the risks associated with the use of AI systems, limiting or prohibiting the use of AI systems that are an unacceptable risk for security – of both citizens and society generally – as well as for people's health, dignity, autonomy, democracy, and other fundamental rights protected on the EU level; and
- Support innovation and excellence in AI systems via the implementation of incentives and funding aimed at developing safe and ethical AI systems, fostering cooperation and coordination among member states, their institutions and all actors and stakeholders involved.

To that end, it is essential that AI systems are regulated based on the level of risk they entail. What does this mean? Basically, the risk assessment of a given AI system evaluates the likelihood that this AI system might bring threats and cause harm to citizens and society. The rationale is that the higher the risk for a given AI category, the more demanding are obligations for providers to follow when using that AI system. There are several categories of risk that are useful to recall:

- Unacceptable risk: there is an explicit ban on the use of AI systems and applications posing unacceptable risks. Into this category fall AI systems capable of manipulating human behaviour, biometric identification in public domains, social scoring (a practice that is sadly used in several countries, especially in Asia, implying the ranking of single persons according to personal characteristics and features, or their economic status, social status, behaviours etc.). Generally speaking, unacceptable AI systems are those capable of interfering in an undesirably acceptably invasive manner on human beings and their fundamental rights;
- High risk: this category includes all AI applications that might pose significant (but not unacceptable) risks for citizens' health, safety, security or any other fundamental right – especially in the domains of education, health, labour law, justice and law enforcement, security management, infrastructural management, and so on. Providers of AI high-risk systems must follow certain steps before putting such systems into the market;

namely, quality and safety assessment, obligations of transparency, and general evaluation of their impact assessment throughout their whole life cycle. Moreover, the AI Act provides a specific list of this typology of systems that can be expanded over time on the basis of new systems potentially causing the same level of risk;

- Limited risk: this category includes AI systems capable of generating contents such as images, audio-visuals, texts, sounds etc. – let us consider, for instance, deepfakes. The limited risk that such AI systems pose entails that the only obligations relate to transparency, which means providers must inform users that such an AI system has specific features; further that users' informed choices are required. The exception is for open source models that, given their public availability and thus greater likelihood of public control, are not regulated because their parameters are, as stated, publicly available;
- Minimal risk: this is mostly associated with gaming-related AI systems or spam filters. There is no regulation provided for this category and, recalling the freedom of their use, implementation and deployment, member states cannot impose local rules forbidding, restricting or limiting their use – in respect of the principles of the EU free market. In case member states do so, the regulation would not apply as it would be considered to infringe upon the EU's hierarchically superior regulation;
- General-purpose AI: this category includes foundation models (e.g., ChatGPT) and is subject to a transparency requirement. It is a hybrid category since general-purpose AI systems might either fall within limited or minimal risk systems (as most of them do) but also be potential high-risk when it comes to systemic risks. Therefore, a case-by-case assessment is required and, although the general rule is simply to undergo a transparency obligation, there are cases in which an in-depth evaluation process might be required;
- Exemptions from risk classification: recital 31 bans "AI systems providing social scoring of natural persons by public or private actors", unless such personal evaluations are carried out for scopes provided by EU law (for instance, tax purposes) – and member states' law accordingly, whether or not infringing on EU regulation; Article 2 of the AI Act expressly excludes from the risk assessment and classification all those AI systems used for military or national security purposes, scientific research and development purposes; Article 5 bans algorithmic video-surveillance when this is conducted synchronously as a general rule, albeit this is possible when the surveillance is used to prevent significant harm and danger such as a real and present, or real and foreseeable, threat of a terrorist attack.

It is worth stressing that most AI systems currently available in the market (or at least largely used by the general public) fall in the categories of limited, minimal or general-purpose AI. High-risk systems are less likely to be used by ordinary citizens, but are instead often used by state or governmental institutions – let us consider the judiciary, health, labour, education, infrastructure etc. If we take a deep look at these systems, they might be compared with the category of special data included in the GDPR when their collection and treatment is justified by a public or relevant interest carried out by public bodies – or by private actors with the backing and authorisation of public regulations. AI systems posing unacceptable risks are less frequent, and the highest risk associated with them is that private actors may use the data collected via such systems (e.g., facial recognition) to disseminate or sell the data for marketing purposes. Still, users' consent in this case does not 'save' such systems due to the fact that the possible harms caused by unacceptable risk AI systems infringe upon citizens' fundamental rights – and, legally speaking, citizens' fundamental rights are not even in the control of the citizen itself, so to say. In this case, the institutions serve as a guardian/protector of citizens' rights and, therefore, citizens' positive actions are not even entailed. Finally, the rationale of special exemptions is clear: with regard to national security, military, research and development, we must consider that a State has a legitimate interest in carrying out such scopes to preserve its integrity and, accordingly, AI Act regulations do not apply since there is no significant threat to citizens.

We have now briefly recalled the main aspects of the AI Act, which was an essential step before comparing its provisions with the Italian draft law on artificial intelligence systems.

The first aspect worth noting is that (as mentioned above) the two regulations under study do not overlap: while the AI Act is mostly focused on assessing the risks a given AI system is capable of bringing to the safety and security of citizens, the draft law aims at regulating the use of AI in several sectors, mainly stressing how AI systems will impact health, education, public administration, judiciary, labour law, cybersecurity, criminal law and intellectual property. Delving deeper into the Italian proposed law, we can observe how specific laws and regulations (in terms of obligations and/or prohibitions) are basically absent while, on the contrary, the AI Act is primarily focused on obligations and/or prohibitions. In this sense, we can state that the draft law appears to be more generic than the EU AI Act, which instead presents regulatory features – also in terms of sanctioning when relevant parties do not follow regulations.

Rather than pointing to the absence of regulatory overlapping, we can say they are actually harmonised, and that this a clear intention of the Italian policymakers aware that when such harmonisation is not present the consequence would basically be the non-application of the domestic regulations.

After further comparison, we notice that the AI Act does not include specific regulations in the sectors of health, labour law, or criminal law, unlike the Italian proposed law. While at first glance this could appear to be a regulatory gap, it is instead justified by the fact that the European institutions only hold limited competences in such fields, that, under the EU Treaties, are mostly member states' exclusive or concurrent/shared competences and hence the EU can merely provide general principles in order to underline regulatory priorities rather than legislate in those sectors. Thus, for instance, the EU can set the goal of tackling cybercrime and call for domestic legislations and sanctions for perpetrators accordingly, but cannot prescribe a specific crime nor the associated sentence. In fact, this is exactly what happens with the use of AI in the judiciary: the EU, namely, the EU Commission for the Efficiency of Justice, recalls as a general principle the need for AI in the judiciary to be regulated given its potentially relevant impact on democracy, the rule of law, and the right to a fair trial, also recommending that AI systems do not interfere with the judiciary. Accordingly, the Italian draft law more specifically regulates the topic by clearly setting boundaries for the use of AI systems in the judiciary.

When it comes to intellectual property and copyright law, the AI Act does not contain specific provisions, while the Italian bill does. The reason for this difference lies in the circumstances that, first, the EU has already implemented a regulatory framework in the field, as detailed as European competences in the field allow: that is, the framework for copyright and neighbouring rights (acquis) consists of 13 Directives and 2 Regulations, resulting in relatively specific provisions in the field, also considering the international treaties (notably WIPO) enforced; second, although member states must follow the EU's general principles in the field, they also have room to regulate certain aspects of IP and copyright (let us think, for example, about the different periods of time that have to elapse before a given work is no longer considered covered by copyright). In conclusion, the proposed law harmonises the provisions of the EU general framework in the field without infringing on the AI Act's regulations – namely, provisions on IP and copyright are basically absent.

In the two acts under examination, we can detect the lack of regulation in the fields of privacy and data protection – albeit, to be fair, the need for AI systems to always ensure compliance with the general principles of data protection is always recalled. Once again, the absence of specific regulation is not an actual regulatory gap. In fact, the whole European framework for the sector is well regulated and covered by the GDPR or the Data Act, that are also applicable in the case of AI systems. Accordingly, the Italian draft law also avoids further regulation that would risk overlap and create confusion.

A final consideration in this comparative analysis: while examining the existing regulatory frameworks we should not forget that, although the AI Act is revolutionary regulation in the field of AI systems, with no equivalents in other parts of the world (e.g., the USA lacks federal regulation in the area of AI), it is not the only piece of legislation to regulate artificial intelligence. The EU framework consists of other acts, such as the abovementioned Digital Services Act, Digital Governance Act or Digital Markets Act (just to mention a few recent ones) that, even though they do not refer specifically to AI systems, are applicable in the field of AI. Of course, none of these acts classify the levels of risk (although the DSA somewhat classifies companies according to potential risks arising from their activities), but at the same time they provide an extensive framework capable of regulating the potential threats deriving from digital devices and their uses. Therefore, also when assessing the Italian draft law we must consider that not only has the proposed law to be harmonised with the AI Act, but it also has to comply with the existing framework – and, despite in-depth analysis of this further comparison falling outside the scope of this paper – we can state that it does. Or, at least, we cannot state that it does not.

3.1 Existing Regulatory Gaps

Notwithstanding the efforts on both the EU side and the Italian side to provide in-depth regulation of AI systems, regulatory gaps still exist. Of course, such gaps have to be tailored to their competences (e.g., in the EU perspective, we cannot view the lack of specific crimes as a gap given member states' competence in the field). Let us consider just a few on both sides.

When it comes to the EU framework:

- The AI Act leaves perhaps too much room for companies' self-assessment with respect to high-risk systems – this poses risks for the transparency and security of a system;
- The AI Act basically considers all of the AI systems used in everyday life (e.g., ChatGPT) as limited or minimal risk systems, and they are thus basically left unregulated – this poses risks for the potential misuse and/or illegal use of such systems without any accountability;
- The AI Act does not sufficiently protect the rule of law and prioritises company and industrial interests – this poses risks for the abuse of citizens' fundamental rights (European Civic Forum, 2024);
- The EU framework is still unregulated adequately in terms of decentralised

platforms: the DSA, the alleged leading regulation in the field of content removal and relative accountability, seems to forget about Web3, AI does not mention it, and only the MICA seems to cover some aspects of the decentralised web and the blockchain, yet the flaw lies in the fact that the MICA is fully finance-oriented – the future of the web is decentralised. Regulatory gaps in the field might be catastrophic; and

- The EU framework is still unregulated when it comes to certain emerging technologies (e.g., the metaverse and mixed reality) that, in a few years, might become game-changers (Interpol, 2024): let us think about virtual social networks and the time that people can spend daily with headsets or a sensor, and this seems to be a growing trend according to several studies in the field – regulations of emerging technologies should be implemented speedily so as to prevent potential harm deriving from the unregulated use of such technologies.

With regard to the Italian framework:

- The draft law lacks clarification when it comes to specific rules regarding the applicability of AI systems. It represents an interesting and (quite) exhaustive overview, but in terms of enforcement there still are regulatory gaps – this brings risks in terms of legal certainty; and
- Although the proposed law is 'brave' enough to criminalise certain conduct facilitated with the use of AI, it is perhaps not 'too brave'. It will be difficult for prosecutors and judges to understand when AI is actually used and to what extent – this poses risks in terms of impunity.

Of course, the abovementioned examples are only some of the gaps that should be addressed to prevent potential threats from becoming actual threats but, before concluding this contribution, one last gap relevant to both the EU and the member states is absolutely worth highlighting. Noting the following considerations (Geopop, 2024; International Energy Agency, 2024):

- ChatGPT consumes over 500,000 kWh in a day (basically the average consumption of 17,200 households);
- In 2027, the AI sector will have an energy demand as high as between 85 and 134 TWh per year: the same as the whole of Ukraine (85 TWh), the Netherlands (108 Twh), Sweden (125 TWh) or Argentina (134 TWh);
- At Google, LLM training brought about a 20% rise of water consumption; at Bing, a 34% rise; the ChatGPT-4 training test led to a 6% rise in the whole of Des Moines;
- AI could consume between 4.2 and 6.6 cubic metres of water in 2027, as much as the entire supply for the United Kingdom;

- One single Meta data centre consumes as much energy as 7 million computers used for 8 hours a day, every day for one year;
- Submitting between 20 and 50 questions to ChatGPT consumes 0.5 litre of water; and
- ChatGPT training creates CO2 emissions equivalent to as much as 125 outbound/inbound Milan/Jakarta flights.

The only conclusion we can confidently draw on this issue is: the European Union urges regulations to reduce the energy consumption associated with the use of artificial intelligence systems. This is not an 'if', not even a 'when', but it should just be a 'now'.

4 Conclusion

In this contribution, we explored the current situation related to the regulation of AI systems in Italy. Although the regulation is a draft law (therefore, a bill – in Italian 'Disegno di Legge'), the likelihood that the document will pass through the legislative iter is high and, thus, in a few months' time the Italian legal framework is likely to include this ongoing regulation. We have presented the primary contents of the proposed law, deepening its fields of application (health, labour law, public administration, judiciary cybersecurity, criminal law, IP and copyright law) and its central outcomes.

Further, we presented the main principles and provisions underlying the EU AI Act, stressing its field of application. Although a thorough assessment of the AI Act would require a separate (and indeed long) analysis, for the purposes of this paper it is sufficient to stress its main contents.

We then compared the AI Act with the Italian draft law, identifying differences and points of similarity, emphasising how, after deeper examination, the two regulations appear harmonised, complementary and well combined to provide nearly exhaustive regulation of AI systems – at least in Italy where the draft law will soon come into force. It appears that there are not many overlapping provisions, which would create critical issues in terms of applicability of the Italian framework, and in the event of any overlapping rules they would not comply with the European rules.

Finally, we pointed out potential regulatory gaps on both the EU and Italian sides (and the two combined) that, if not addressed, could raise more than one issue with respect to the implementation of AI systems in society, especially whether such systems will be used on a daily basis by ordinary users – e.g., social networks and their components.

We summarised the findings in the table below, showing the legal fields regulated by the draft law or the AI Act.

Table 1: Regulations in different legal fields

Legal field	Italian draft law	EU AI Act
Labour Law	×	N/A
Health and disability	×	N/A
Public Administration	×	N/A
Judiciary	×	N/A
Cybersecurity	×	×
IP and copyright	×	N/A
Criminal law	×	N/A
Accountability	N/A	×
Obligations for stakeholders	N/A	×
Ethics	×	×
Democracy	×	×
Citizens' fundamental rights	×	×
Environmental law	N/A	N/A

Source: CNIL

In order to overcome these regulatory gaps, it is worth underlining the following recommendations addressed to policy- and law-makers that, if addressed, would help make AI a trustworthy source of human development:

a. Amendments to (or at least explanatory interpretations of) the AI Act are required in terms of companies' self-assessment of high-risk AI systems and the lack of regulation for limited or minimal risk AI systems;

b. On a general level (except for cryptocurrencies), the current EU legal framework is lacking regulations that specifically address legal conundrums arising from the use of decentralised internet architectures, especially in social media platforms whose users are generally ordinary citizens (e.g., content moderation above all);

c. The metaverse, or virtual worlds, or mixed reality worlds, or however one wants to call it, is a growing phenomenon. Let us not be confused by the decrease in the hype surrounding it: the metaverse is there and will be the future of social networks. There are presently no regulations at all addressing it, and this is an urgent need – especially when it comes to crimes (mostly, sexual harassment, as several police reports have shown), and IP and copyright law; and

d. In times in which the environment, its sustainability, and its protection appear to be a priority for the EU and governments around the world, it is quite curious that the issue of the amount of energy consumption deployed by AI systems – as well as cryptocurrency – is not addressed at all by the EU institutions. Regulations are required at least to limit such consumption and introduce boundaries around the possibly unlimited increase of it. This is perhaps the most pivotal and urgent need above all others.

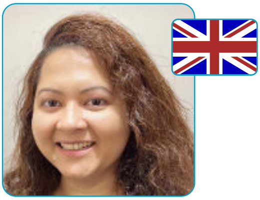
In conclusion, while the general regulatory legal framework appears adequate for addressing issues arising hitherto from the use of AI systems, the path ahead is still long. The EU and member states should combine efforts to adopt comprehensive and harmonised policies to make AI exclusively an opportunities-generator, not a threat-generator.

REFERENCES

- Agenda Digitale. (2024) AI Act: cos'è e come plasma l'intelligenza artificiale in Europa. Retrieved 10 July 2024 from: https://www.geopop.it/nei-prossimi-anni-lai-potrebbe-consumare-tanta-energia-quanto-unintera-nazione
- Consiglio dei Ministri. (2024) Comunicato stampa n. 78 del 23 aprile 2024. Retrieved 31 July from https://www.governo.it/it/articolo/comunicato-stampa-del-consiglio-dei-ministri-n-78/25501
- European Civic Forum. (2024). Packed with loopholes. Why the AI Act fails to protect civil space and the rule of law. Retrieved 7 July 2024 from: https://civic-forum.eu/advocacy/artificial-intelligence/packed-with-loopholes-why-the-ai-act-fails-to-protect-civic-space-and-the-rule-of-law
- European Commission. (2024) AI Act. Retrieved 4 July 2024 from: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- Geopop. (2024). Nei prossimi anni l'AI potrebbe consumare quanto un'intera nazione. Retrieved 11 July 2024 from: https://www.geopop.it/nei-prossimi-anni-lai-potrebbe-consumare-tanta-energia-quanto-unintera-nazione
- International Energy Agency. (2024). Retrieved on 8 July from: https://www.iea.org/
- Interpol. (2024). Metaverse: A Law enforcement perspective.
- Nature. (2024). There are holes in Europe's AI Act – and researchers can help to fill them. Retrieved 12 July 2024 from: https://www.nature.com/articles/d41586-024-00029-4
- Senato della Repubblica. (2024). Disegno di Legge. Disposizioni e delega al governo in materia di intelligenza artificiale. Atti parlamentari. Senato della Repubblica. N. 1146. Retrieved 10 July 2024 from https://www.senato.it/service/PDF/PDFServer/DF/437373.pdf
- Sorrentino, C. (2024). Intelligenza artificiale: cosa prevede la normativa italiana. Osservatori.net Digital Innovation. 6 May 2024. Retrieved 5 July 2024 from: https://blog.osservatori.net/it_it/intelligenza-artificiale-cosa-prevede-normativa-italiana

Chapter 4

Regulation for Systemic Artificial Intelligence Safety: Interoperable Regulations and a Pragmatic Enforcement Strategy for Mitigating Unintended Consequences



Farhana Ferdousi Liza

1 Introduction

Developing and managing systemically safe, advanced Artificial Intelligence (AI) technologies has become increasingly critical due to the rapid development and deployment of AI systems across various sectors, together with the evolving regulatory landscape. Ensuring AI safety, trustworthiness and fostering global cooperation are paramount for effective governance and maintaining power relations conducive to a liberal future. Achieving these goals calls for a balanced regulatory approach and being prepared for unintended consequences. A systemic and pragmatic approach to AI regulation can assure the safe, responsible use of advanced AI technologies while promoting cross-border business and technological research and innovation.

Governments and regulatory bodies around the world are developing regulations to address a range of AI-related concerns like safety, ethics, bias, fairness, security, privacy, transparency, and accountability. Notable examples of such regulations include the European Union's AI Act (EU AI Act), the United States' AI Bill of Rights, and various national AI strategies, including the United

Kingdom's (UK) National AI Strategy. Stanford University has reported a huge increase in the number of countries with laws containing term "AI" – growing from 25 countries in 2022 to 127 in 2023 (Marcin, 2024). These regulations aim to protect rights and liberties by safeguarding individuals' rights and preventing discrimination. They envision compliance by establishing standards and guidelines for the development and deployment of AI systems, and promote innovation by encouraging the responsible development of AI technologies that benefit society and the economy.

While regulations are essential for guiding the safe and legal use of AI, it is crucial to analyse their unintended consequences, particularly perverse results, to make sure they do not stifle innovation or create new problems. Overly stringent, partially analysed and difficult-to-enforce regulations might hinder technological innovation, increase discrimination, and slow the advancement of beneficial AI research and innovation. Regulations can hold considerable social, economic, and power implications, affecting businesses' operating costs and competitiveness. Understanding these impacts is vital for designing policies that support economic growth while ensuring responsible AI use in a liberal society.

Small and medium-sized enterprises (SMEs) and the public sector may find it challenging to comply with complex regulations, potentially leading to less competition and innovation and making large enterprises more powerful. Assessing the compliance burden can guide the creation of more accessible and scalable SME friendly regulatory frameworks. Different regions are adopting varying regulatory approaches, bringing inconsistencies and potential disparities in AI development and deployment globally. Analysing these differences can promote harmonisation and international cooperation. Some regulatory requirements may be technologically challenging to implement. Analysing the feasibility of these requirements assures they are practical and achievable with the current knowledge of science and technologies. Further, regulations can influence the behaviour of developers, users and organisations. Understanding these incentives can help predict and mitigate any negative behavioural changes, such as gaming the system, finding loopholes, or focusing on compliance over innovation.

This chapter analyses issues in international cooperation (interoperability) and technological challenges in implementing AI technologies able to comply with current AI regulations across borders. The discussion aims to foster responsible AI development through a systemic approach to safety, a forward-looking AI regulatory system and a pragmatic enforcement plan within the European Union (EU) and beyond.

We first discuss the unintended consequences of past regulations and technology adoptions as case studies in Section 2. We show that striking the right balance in regulation across key parameters – timing, scope and method – is indispensable for mitigating unintended consequences. Next, in Section 3 we conduct a concise evidence synthesis to gain an understanding of the global landscape of AI regulation. We then briefly study the regulatory goals of key jurisdictions – the UK, the USA and the EU – in Section 4. In Section 5, we examine cross-cutting challenges in enforcing AI regulation across borders. Section 6 explores divergent regulatory approaches to AI. In Section 7, we address the problem of interoperability between statutory and non-statutory regulatory frameworks. Section 8 presents a computational analysis of the enforceability of data requirements as an example. Finally, before concluding, in Section 9 we combine our analysis of AI regulation and unintended consequences with the framework provided by the sociologist Robert K. Merton.

2 Unintended Consequence: Drawing Insights from Historical Technology Adoption and Regulations

In regulatory analysis and debate, anecdotal evidence and case studies offer valuable insights. Studies in tissue engineering (Faulkner, 2009), agro and pharmaceutical biotechnology (Chataway et al., 2006) and pharmaceuticals (Abraham and Davis, 2007) reveal the scientific and pragmatic value of detailed analyses often grounded in social or political theory. Given the evolving nature of AI regulations, it may be premature to draw definitive theoretical conclusions. Still, historical examples can illuminate the main parameters that guide our analysis concerning the unintended consequences of AI regulations. The emergence of new technologies and their adoption, coupled with the introduction of new regulations, is not unprecedented. Examining historical examples helps us understand the impacts of regulations and technology adoption on people, the economy, society, and businesses over time.

Example 1: Impact of strict regulation on economic growth

An illustrative case of how strict regulation can hinder economic growth is India's Licence Raj system that was in place from around 1950 to 1991. The Licence Raj comprised the elaborate system of licences, regulations and red tape required to set up and run a business in India. These restrictive policies caused sluggish economic growth, often referred to as the "Hindu rate of growth", which hovered around 3.5% annually. The economy remained largely agrarian, with slow industrial development and high levels of poverty.

Economic reforms in 1991, which included reducing government control over businesses, lowering tariffs, and encouraging foreign investment, led to a significant acceleration of the economy. India emerged as one of the fastest-growing major economies in the world. Declassified documents from the World Bank and International Monetary Fund (IMF), reported by The Indian Express in 2010, reveal how both organisations had pressed India to transition from a centrally planned to a market-driven economy. This example shows the potential of overly stringent regulations to stifle economic growth and innovation.

The EU's AI Act is considered to be a strict regulation (Euronews, 2023). Although it is challenging to foresee all the unintended consequences of this regulation, given that it only came into effect on August 1, 2024, historical examples and existing literature suggest that stringent regulations generally inhibit research and innovation (Stewart, 1981).

Example 2: Impact of faulty technology on people and society

In 1999, while the UK was still a member state of the EU, the Postmaster Scandal, also known as the Horizon scandal, unfolded. It entailed the wrongful prosecution of over 700 sub-postmasters and sub-postmistresses due to errors in information technology (IT) support for UK post offices – the Post Office's Horizon IT system. Despite early indications of discrepancies, the Post Office steadfastly defended the system's reliability, resulting in numerous unjust convictions for theft and fraud. The lack of effective central oversight compounded sector-specific issues. Ideally, the system should have undergone a thorough evaluation both before and during its deployment to mitigate the technical loopholes and perceived vulnerabilities. Persistent advocacy efforts and a pivotal 2019 High Court ruling finally exposed the systemic flaws. The scandal was responsible for serious consequences, including financial ruin, mental health challenges and, tragically, suicides among those affected. In response, the UK government initiated a compensation programme and launched a public inquiry to scrutinise the shortcomings. This case points to the critical need for stringent but pragmatic and enforceable technological oversight (both centrally and institutionally), transparency, and comprehensive institutional reforms.

Example 3: Impact of a regulatory loophole or hasty regulation on people and society

It is often not fully appreciated that the Enabling Act of 1933 in Germany was not an unheard of measure within its historical context. This legislation granted Adolf Hitler's government the power to enact laws without needing to involve the Reichstag, effectively establishing a dictatorial authority. Prior to 1933, the unstable Weimar Republic had already experienced a series of enabling acts intended to circumvent constitutional challenges. This historical example highlights the importance of considering the broader implications of such hasty measures – addressing one problem may inadvertently create another, potentially more pressing issue. Moreover, loopholes in the legal system and the influence of money can undermine regulatory frameworks. For instance, in the USA, Robert Durst (Editors, 2022), despite admitting to having 'accidently' killed and dismembered a body in 'self-defence', was able to avoid jail time by being able to recruit the best and most expensive lawyers and leveraging substantial bail funds, thereby circumventing both legal consequences and social stigma associated with his actions.

Although these analogical comparisons are disturbing given their association with mass genocide and killing, they may be useful for understanding the scenario. In the contemporary context of regulating artificial intelligence (AI), one should be mindful of this lesson – the hasty making of regulations, underestimating the power of money, and the inevitability of legal loopholes. While regulation is necessary to manage the risks associated with AI, care must be taken to avoid introducing new problems that could arise from cross-boundary business needs and the imperative for research and innovation. Balancing regulation with the need to foster technological advancement requires careful and thoughtful policymaking to make sure the solution (regulation) does not become a bigger issue than the original problem.

On a hasty regulation note, some European legislation on AI had been anticipated at least since 16 July 2019. On that date, Ursula von der Leyen pledged that, within 100 days of being elected President of the European Commission, she would propose new legislation on AI (Floridi, 2021). It remains unclear, though, how this 100-day timeline was determined.

Example 4: Impact of late technology adoption on businesses

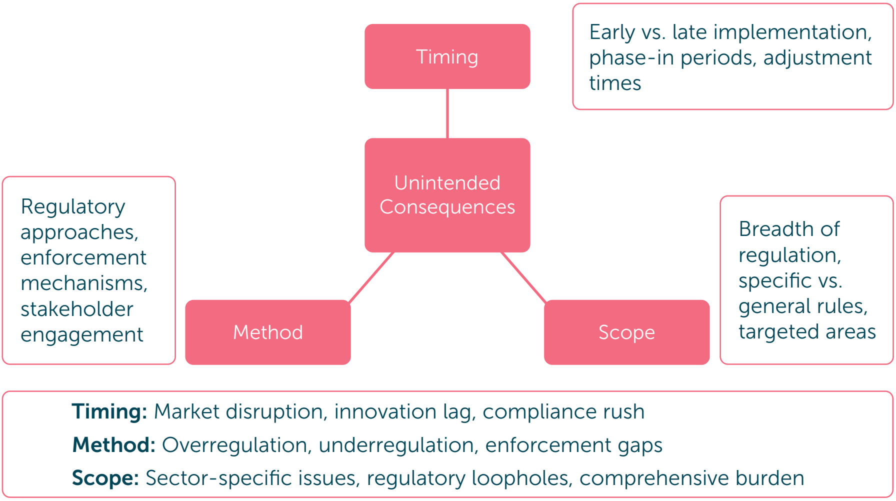
Kodak's decline following to delay in adopting digital photography serves as a poignant example of the adverse consequences of postponing technological innovation. Despite pioneering the first digital camera in 1975, Kodak remained focused on its lucrative film business, thereby missing the digital revolution led by competitors like Canon and Sony. This strategic hesitancy led to a significant loss of market share, financial deterioration and, ultimately, bankruptcy in 2012. Kodak's downfall shows the perils associated with resistance to innovative changes and misalignment with prevailing market trends. While an individual entity's adoption may impact its trajectory (e.g., Kodak), regulations can impact a whole region systemically. The systematic adoption of new technologies can be complicated by stringent regulations (e.g., the EU AI Act), added compliance burdens, and heightened liability (Wendehorst, 2020). Such regulatory challenges hold the potential to stifle innovation as they create additional obstacles for businesses attempting to adapt to and integrate emerging AI technologies.

Striking the right balance in key parameters (timing, scope, method)

From these examples, we may conclude that achieving the right balance in adopting technology and regulating those technologies is vital. Based on the above case-study discussion, to mitigate unintended consequences, three main parameters must be considered: timing, method and scope. First, appropriate timing involves implementing regulations neither too early nor too late, with phase-in periods and adjustment times to facilitate smooth adoption. These strategic tools help stakeholders ease into compliance, reduce immediate burdens, and allow for necessary infrastructure and skill development. For instance, the GDPR's 2-year transition period and new vehicle emission standards' phase-in periods illustrate this approach. Second, the method must be stable, specific and future-proof so as to enable flexible enforcement and mitigate unintended consequences. This includes making sure of the interoperability of regulatory approaches and robust enforcement mechanisms, as well as appropriate stakeholder engagement. Finally, the scope involves determining what to regulate, the breadth of regulation, and whether to use specific or general rules, focusing on targeted areas. For AI, this is particularly complex due to cross-cutting challenges, which means it is essential to carefully balance these parameters to support long-term compliance, research and innovation. The figure below provides an illustration.

> ”
>
> Achieving the right balance in adopting technology and regulating those technologies is vital.

Figure 1: Striking the right balance in key parameters (timing, scope, method)


Timing
Early vs. late implementation, phase-in periods, adjustment times
Unintended Consequences
Regulatory approaches, enforcement mechanisms, stakeholder engagement
Method
Scope
Breadth of regulation, specific vs. general rules, targeted areas

Timing: Market disruption, innovation lag, compliance rush

Method: Overregulation, underregulation, enforcement gaps

Scope: Sector-specific issues, regulatory loopholes, comprehensive burden

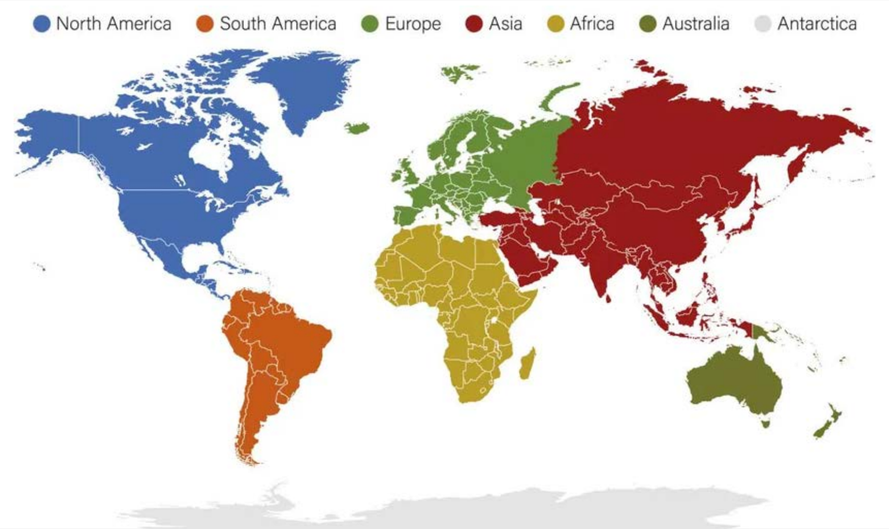
Source: Own

3 The Evidence Synthesis of Global AI Regulations

This section provides a concise overview of how AI regulation is developing around the world. The rapid advancement of AI is prompting countries and regions to develop regulations tailored to their unique cultural and political landscapes. According to the September 2023 Global AI Legislation Tracker, "countries worldwide are designing and implementing AI governance legislation commensurate to the velocity and variety of proliferating AI-powered technologies. Legislative efforts include the development of comprehensive legislation, focused legislation for specific use cases, and voluntary guidelines and standards" (Marcin, 2024).

Continent-wise (see Figure 2), in North America the United States is shaping its regulatory landscape through federal initiatives and ethical principles along with state-based legislations, whereas Canada is advancing its AI oversight with the Pan-Canadian AI Strategy and Algorithmic Impact Assessment (AIA). In Asia, China's National AI Development Plan and Ethical Guidelines, Japan's AI Strategy 2021 and contributions to global AI standards, and other initiatives are driving regulatory developments. Europe is considerably influenced by the EU's AI Act, with individual countries like the UK, Germany, France and Italy also pursuing their national AI strategies. In South America, Brazil is at the forefront, while in Africa, South Africa and Kenya are leading in AI strategy and regulatory development. Australia complements its AI Ethics Framework and AI Action Plan with regulatory sandboxes to test AI applications in controlled conditions. Notably, Antarctica remains without any dedicated AI regulatory initiatives.

Figure 2: Continents of the world



Source: Toby, 2023

Despite the diversity of these regulatory efforts, common themes emerge, particularly in addressing ethical and legal considerations to ensure safety, security, transparency, fairness, accountability, and respect for human rights. Ethical frameworks often precede legal regulations, given their foundational role and slower legal development pace, especially when interoperability is critical and International Law is suggested as the principal legal framework for the regulation of AI (Carrillo, 2020). Promoting AI research, development and innovation via public-private partnerships, funding and international collaboration is another common priority, albeit the emphasis varies by region. The balance between promoting innovation and ensuring regulation also differs, with the EU generally adopting more stringent controls than the more flexible, innovation-friendly approaches of the USA and UK. Countries like the USA, the UK and Australia favour sectoral regulations, addressing specific industries such as health, transportation, finance, the public sector, smart city technology, and robotics. In contrast, the EU is aiming for a comprehensive, risk-classified cross-sectoral framework. Some regulations are statutory, as seen in the EU, whereas others are non-statutory with a possibility of statutory regulation in the future, such as in the UK, which would have long-term implications for adoption and enforcement.

Given the substantial early contributions to frontier AI developments from USA industries, it is likely that the USA will play a leading role in setting international standards for AI regulations. The preference for sectoral regulation, which is currently favoured by several countries and in line with the USA approach, contrasts with the EU's risk-based methodology. The entity that establishes the dominant and pragmatic regulatory framework for AI and possesses the capacity and influence to enforce it will significantly impact the global balance of power concerning AI.

4 Goal of the AI Technology Regulation by Key Jurisdictions

In this section, we are considering the goals and approaches of key jurisdictions such as the EU, USA and UK, which are continuously working to regulate AI. Analysing all jurisdictions lies beyond the scope of this chapter. The EU has finalised legislations to regulate AI which came into effect on 1st August 2024. In contrast, the Conservative UK government argued that it was premature to legislate effectively given the present stage of AI technology's evolution, suggesting that doing so now might be counterproductive. The Artificial Intelligence (Regulation) Bill, a private member's bill proposed by Lord Holmes of Richmond, seeks to establish a new body, the AI Authority, tasked with addressing AI regulation in the UK. Following the UK General Election on 4 July 2024, the new Labour government may well change the current approach (BTO, 2024). In the USA, AI regulation is being explored on both federal and state levels. While some states have introduced legislation focusing on privacy and accountability, there is no federal legislation yet. Instead, the White House issued an executive order in October 2023 outlining key principles and actions for the safe development and use of AI. Other countries are also initiating AI regulations, each at different stages of development.

Given the EU's advanced stage in regulating AI, the purpose of the regulation is articulated in Article 6 of the EU's AI Act: "The purpose of this Regulation is to improve the functioning of the internal market and promote the uptake of human-centric and trustworthy artificial intelligence (AI), while ensuring a high level of protection of health, safety, and fundamental rights enshrined in the Charter, including democracy, the rule of law, and environmental protection, against the harmful effects of AI systems in the Union and supporting innovation". There are 85 articles, and several governing bodies are set up for proper enforcement, including an AI Office within the Commission to enforce the common rules across the EU, a scientific panel of independent experts to support enforcement activities, an AI Board with member states' representatives to advise and assist the Commission and member states on consistent and effective application of the AI Act, and an advisory forum for stakeholders to provide technical expertise to the AI Board and the Commission.

In contrast, the UK government outlined its approach to AI regulation in the March 2023 White Paper "A Pro-Innovation Approach to AI Regulation". The White Paper does not propose creating a new AI regulator or introducing primary legislation to establish AI regulatory principles or structures. Instead, the government intends to implement a new framework to bring "clarity and coherence" to the AI regulatory landscape, underpinned by five principles: safety, security and robustness; appropriate transparency and explainability; fairness; accountability and governance; and contestability and redress. These principles will be addressed on a non-statutory basis and implemented by existing sectoral regulators. The framework relies on the UK's existing sectoral regulators, supplemented by government-led "central functions" to provide support, coordination and coherence. Therefore, the AI regulation approach is still in progress but remains pro-innovation and the UK government has allocated GBP 10 million in funding for regulators to develop skills and enhance capabilities.

In the United States, AI regulation is being examined on both state and federal levels. On the federal one, the White House issued an executive

> The UK government intends to implement a new framework to bring "clarity and coherence" to the AI regulatory landscape.

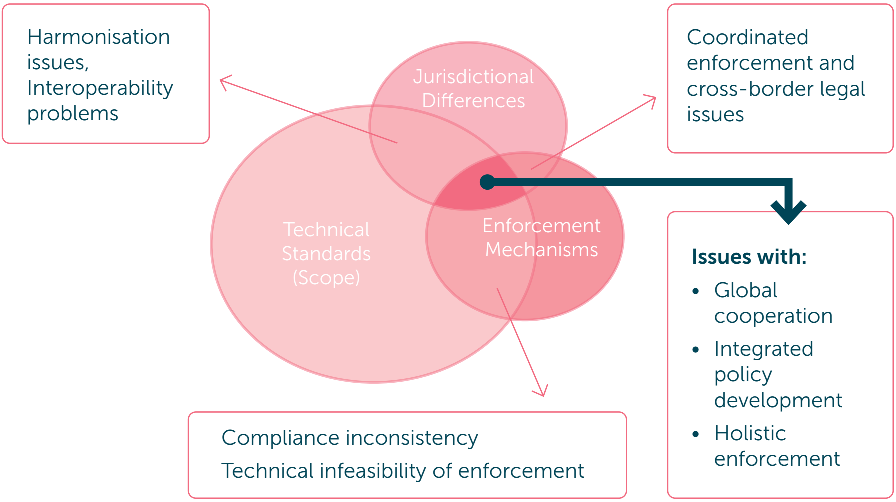
order in October 2023 directing new standards for AI safety and security and several measures on privacy. The order also contained measures on advancing equity and protecting civil rights; the impact of AI on consumers, patients and students; supporting workers; promoting innovation and competition; assuring responsible and effective government use of AI; and "advancing American leadership abroad". Under the last heading, the order refers to the expansion of bilateral, multilateral and multi-stakeholder engagement, and the acceleration, development and implementation of "vital AI standards with international partners and in standards organisations, ensuring that the technology is safe, secure, trustworthy, and interoperable". Since 2019, 17 states have enacted 29 bills focused on regulating the design, development and use of AI. These bills primarily address two regulatory concerns: data privacy and accountability. Legislatures in California, Colorado and Virginia have established regulatory and compliance frameworks for AI systems. In the USA budget announced in March 2024, President Biden also noted there would be significant new funding (over USD 3 billion) aimed at developing responsible AI.

While key jurisdictions take distinct approaches to AI regulation, the USA aims to promote interoperability with international regulatory frameworks. The chief goal of regulating AI across different jurisdictions is to make sure that AI technologies are developed and utilised in a way that maximises their benefits for innovation and the economy, while mitigating potential risks and harms to individuals, society and the environment. The implementation of AI regulation goals varies in several respects: timing (e.g., the EU has already implemented regulations, whereas the UK and USA are still in progress), method (e.g., the EU adopts a strict but innovation-supportive approach, while the UK and USA are more pro-innovation, with the USA focusing on cross-border interoperability and the EU on regional interoperability), and scope (e.g., the EU employs a risk-based approach, while the USA and UK use a sectoral approach). Further, budget allocations for enabling responsible AI development and deployment vary, with the UK allocating GBP 10 million and the USA USD 3 billion.

5 Cross-Cutting Challenges with Enforcing AI Regulation Across Borders

Now that we understand the goals of different jurisdictions, we shall turn to how to technically support compliance and enforcement. Shetty (2024) noted that "The EU needs the technical standards supporting its AI Act to be restrictive enough to protect consumers, but flexible enough to enable innovation. Given society's current understanding of AI, there are serious doubts as to whether such standards are technically feasible". Cross-cutting challenges (CC) impact all aspects of regulation development and enforcement plans, influencing each other. We will discuss the major challenges and their relationship to unintended consequences within the scope of this chapter, as illustrated in the figure below.

Figure 3: Cross-Cutting Challenges (in circles) and Unintended Consequences (in boxes) in AI Regulation


Harmonisation issues, Interoperability problems
Jurisdictional Differences
Coordinated enforcement and cross-border legal issues
Technical Standards (Scope)
Enforcement Mechanisms
Issues with:
Global cooperation
Integrated policy development
Holistic enforcement
Compliance inconsistency
Technical infeasibility of enforcement

Source: Own

Cross-cutting challenges 1 (CC1):

Part 1: Technical standard – lack of a standardised definition (related to the Scope parameter)

The lack of public awareness about AI largely stems from insufficient knowledge. Mythology, culture, religion, literature and science fiction have contributed to an anthropomorphic view of AI (Ramírez, 2018; Muehlhauser & Helm, 2012). Technologically compliable effective regulation of artificial intelligence in contrast requires precise and universally accepted terminologies. The fact that currently there is no cross-border consensus on the legal definition of AI (Samoili et al., 2020; Begishev et al., 2020; Schuett, 2023) poses significant challenges for the regulatory clarity and consistency, policy development, trusted AI innovation, and establishment of uniform ethical and safety standards. This fundamental issue complicates the sociotechnical landscape, notably for advanced AI technologies like Large Language Models and innovations such as ChatGPT. Within existing regulatory frameworks, these technologies can be categorised as either high- or low-risk depending on their application – high-risk in healthcare, for example, and low-risk in the music industry.

Autonomy is another key term used in regulations to describe AI characteristics. However, what constitutes autonomy remains ambiguous. Advanced AI technologies capable of learning and producing contextualised outcomes present greater legal and enforcement challenges than purely technical uncontextualised automation, such as vending machines. For example, an AI system is not akin to an automated coffee vending machine where pressing a button brings about a predictable outcome – a cup of coffee. Each AI system varies significantly in its learning capabilities and types of autonomy. For simplicity, autonomy can be conceptualised as the degree of variance with respect to a context inherent in an AI system.

Moreover, regulations often call for the explanation of decision-making processes. But what exactly is meant by explanation? In the sciences, rationalism's focus on reason, logic, and clear deductive processes enhances the explainability of knowledge. Rationalism in social science creates logical, clear and systematic interpretive explanations of social phenomena. Scientific explanations aim to discover universal laws and principles that can predict and explain natural phenomena. In contrast, social science seeks to understand and interpret social phenomena within specific contexts rather than discovering universal laws.

For advanced AI models, providing explanations involves offering understandable reasons for the model's outputs or decisions. While interacting with humans, these explanations need to adapt to specific contexts. AI models, which identify patterns and generate responses based on their training data, require clarity concerning what type of explanation (deductive vs. interpretive) is mandated by AI regulations. This clarity is essential for ensuring that AI systems are transparent and accountable.

Part 2: Technical standard – tackling challenges in logical reasoning, explanation and uncertainty quantification in AI systems (related to the Method parameter)

Artificial Intelligence began to emerge as a field in the mid-20th century, influenced by seminal events such as the proposal of the Turing Test to assess machine intelligence and coining of the term "artificial intelligence" at the 1956 Dartmouth Conference where researchers gathered to discuss how machines could simulate human intelligence. A few decades later, advanced AI technologies like Large Language Models (LLMs) and ChatGPT reached a development stage with the potential to disrupt the socio-legal-technological landscape. These technologies rely on information theory, statistics, and probability theory, utilising evaluation measures such as perplexity and accuracy during training. However, systematic reasoning strategies like logic, decision trees or uncertainty quantification methods such as Bayesian modelling were not technically feasible in the training objectives of these models, which often use algorithms like backpropagation for the neural networks. From the beginning to the present, two main philosophies – rationalism and empiricism – have been explored in the AI domain, specifically in computational linguistics, within which Large Language Models (LLMs) fall. These philosophies can be divided into four eras, as shown in Table 1.

Table 1: Historical stages (Church & Liberman, 2021) of the philosophical adoption of research approaches in AI (specifically Computational Linguistics)

Era	Philosophy	Approach
1950s	Empiricism I	Information theory
1970s	Rationalism I	Formal language theory and logic
1990s	Empiricism II	Stochastic grammars
2010s	Empiricism III	Deep neural networks

In Table 1, logic played a larger role during the rationalism emphasis in the 1970s, while probability was more prominent in the empiricism phases during the 1950s and 1990s. In the 2010s, both logic and probability faded into the background as deep neural networks introduced a procedural, associationist flavour of empiricism. One might ask why rationalism has not been revived until now. Despite numerous attempts to incorporate reasoning into state-of-the-art advanced AI technologies, no significant breakthroughs have been made. Achieving reasoning in advanced AI technologies remains a considerable challenge within the current scope of known science. The integration of logical reasoning and uncertainty quantification into AI systems is essential for advancing their capabilities and assuring robust and trustworthy outcomes, although that requires significant technical hurdles to be overcome. Imposing laws on requirements that are technically infeasible seems to amount to unrealistic criteria in the current landscape of legislations.

Part 3: Technical standard – challenges in establishing a social AI agent align with the EU AI Act (related to the Scope, Timing and Method parameters)

What these regulations are calling for is a Social AI Agent – one capable of understanding ethics, social customs such as fairness, able to adapt the explanation of an activity of decision based on the context, and essentially be part of society. Creating a socially adept AI agent necessitates significant technical and methodological advancements in AI, natural language processing, computer vision, machine learning, and ethical AI design. For example, transforming ChatGPT into a social agent capable of engaging in human-like conversations involves overcoming several key challenges. First, ChatGPT must adapt its responses to diverse cultural contexts and continuously learn from user interactions to handle complex social scenarios effectively. This requires a nuanced understanding of the context and intent behind user queries. Developing emotional intelligence within ChatGPT is also crucial in such scenarios; the AI needs to recognise and respond empathetically to user emotions to facilitate meaningful and safe interactions. In addition, integrating multimodal capabilities to handle inputs such as text, images and videos can enrich the conversational experience. Consistency in dialogue, respect for privacy, and adherence to ethical standards are essential for building and maintaining user trust. Nonetheless, known science and technology has not invented these mechanisms yet and we do not know when that will be possible.

Cross-cutting challenges 2 (CC2):

Jurisdictional differences – innovation across borders and adapting cross-border contracts (related to the Method parameter)

The vagueness of definitions along with technically infeasible criteria in regulations complicates the harmonisation of international regulations on business and sales across various systems and contexts. While one can find existing legal mechanisms like the United Nations Convention on Contracts for the International Sale of Goods (CISG), questions have been raised about whether changes in contract law are needed to accommodate issues specific to AI (Janssen, 2022). The CISG aims to provide a uniform and equitable framework for international trade by harmonising the laws governing cross-border sales transactions.

However, emerging concerns with AI are that existing frameworks like the CISG might not adequately address the unique challenges posed by AI technologies. These challenges include the autonomy of AI systems, which can affect contract performance and liability. Therefore, a critical examination of how current international contract laws, such as the CISG, can be adapted to better address the complexities introduced by AI is essential. This includes considering new definitions and legal standards that can harmonise international AI regulations, thereby promoting businesses and innovation while ensuring legal clarity and fairness across borders.

Cross-cutting challenges 3 (CC3):

Technical standard, jurisdictional differences and enforcement mechanisms – developing fit-for-purpose transdisciplinary research (related to the Scope, Method and Timing parameters)

A pragmatic and enforceable AI regulation can be achieved by integrating sophisticated techniques across formal, natural and empirical sciences (such as logic, physics, mathematics, statistics, and computer science), social sciences (including sociology, psychology, political science, law, and education), and engineering disciplines (e.g., GPU technology). Effective AI regulation calls for a comprehensive understanding that transcends individual scientific and engineering domains. AI systems, designed to operate within a human–computer interaction framework, embody principles from these diverse fields, combining both hardware and software components. Human involvement by way of designers, developers, policymakers and users introduces crucial social science elements into the equation. The call for secure, trustworthy and ethical AI, alongside reqand engineering domains. Technocrats who adopt a transdisciplinary approach – seeing the whole picture – offer a promising path forward to resolving these issues. The interdisciplinary collaboration needed to create and enforce effective regulations adds another layer of complexity, underscoring the fundamentally transdisciplinary nature of the problem of AI regulation.

> ”
> Effective AI regulation calls for a comprehensive understanding that transcends individual scientific and engineering domains.

6 Divergent Regulatory Approaches to AI

“More law, less justice” is a quote by Marcus Tullius Cicero that conveys the idea that too many laws can lead to injustice. We now elaborate on jurisdictional differences, inconsistencies and potential scope for injustice. The regulatory strategies employed by the European Union (EU), the United States (USA) and the United Kingdom (UK) reflect fundamentally divergent methodologies. The EU’s risk-based framework operates on a macro level, utilising top-down regulation, whereas the USA and UK have adopted sectorial approaches that function on a micro level with bottom-up regulation. While these approaches hold the potential to complement each other through rigorous international collaboration, they also bring risks of unintended consequences, particularly during transitions between the sectorial and risk-based frameworks essential for achieving cross-border interoperability.

Challenges in interoperability and consistent enforcement

Interoperability and consistent enforcement pose considerable challenges not only among the EU, UK and USA, but also within the EU itself if any member state wants to adopt a bottom-up approach in the future. Concerns have been already expressed regarding the alignment between the EU AI Act and the regulatory and policy implementations of member states. For instance, Gilbert (2024) notes that “the wording of many aspects of the Act is ambiguous, with high-level objectives stated and details expected in subsequent guidance, standards, and member state laws”. The prevalence of vague objectives and lack of specificity are recurring themes in the literature (Liza, 2022).

Moreover, the EU AI Act imposes new responsibilities on developers, deployers, notified bodies, regulators, and the newly established EU AI Office and European Commission. Its impact extends to AI-driven technologies (e.g., Digital Health Technologies) developed and deployed both within and outside the EU (Gilbert, 2024). Innovation in public sectors like healthcare is consequently likely to decelerate due to inadequate funding for routine operations compounded by vague legislation and heightened compliance burdens. This lack of clarity, combined with regulatory burdens and interoperability issues, is anticipated to hinder innovation.

Exclusions and their Implications

While the legislation imposes stringent regulations on AI-based technologies, its exclusion of military and defence sectors complicates interoperability and introduces the risk of discriminatory practices within risk-based regulations.

European Digital Rights (2024) observes that "the legislation establishes a separate legal framework for AI use by law enforcement, migration control, and national security authorities, potentially creating loopholes and endorsing the use of intrusive systems for discriminatory surveillance on marginalised communities".

As a result, even relatively low-risk AI innovations in public sectors such as healthcare may stall due to concerns with compliance costs, while more controversial AI applications in military and defence could evade scrutiny, potentially exacerbating discrimination against marginalised groups, including migrants. The term "migrant" lacks a universally accepted legal definition in international law and generally refers to individuals who have left their homes, whether within their own country or across borders.

Regulatory coherence and potential self-contradictions

This situation raises significant concerns regarding the coherence of regulations and potential self-contradictions. If migrants from EU member states are subjected to discriminatory AI applications by military and defence sectors of other member states, that would violate their fundamental human rights and contradict the EU AI Act's objective of safeguarding the rights of all EU citizens.

7 Statutory vs. Non-Statutory Regulations

Such jurisdictional differences reveal variations in enforcement plans. Regulatory approaches to AI can be broadly categorised into statutory and non-statutory frameworks. The EU is adopting a statutory regulatory approach; the USA has not decided yet – either as standalone legislation or as AI-related provisions and clauses inserted into broader acts – so at the federal level, the guidelines are non-statutory, and the UK is adopting non-statutory regulation and when the time is right the UK might take a statutory approach. As we can see, the 'timing' parameter is playing its role here. Statutory regulations are formal laws enacted by legislative bodies, whereas non-statutory regulations consist of guidelines, principles, and codes of practice that lack legal enforceability. This section explores the interaction between these two regulatory types and examines the unintended consequences that may arise when 'timing' is assessed differently.

Multiple entities will be subject to AI regulations (both statutory and non-statutory), including small and medium-sized businesses, large companies like Facebook, Google and Twitter, public sector organisations, and the financial sector. If a UK-based entity sells or distributes its AI tools within the EU, those tools must comply with the EU's AI Act. Similarly, if EU-based businesses use AI tools developed in the UK or any other country, those tools must also comply with that Act. Entities willing to grow and adopt AI innovation but struggling with compliance, vagueness and the burden of regulation might consider relocating operations to areas with non-statutory regulations or exploiting regulatory loopholes resulting in compliance on paper but not in spirit, and potentially creating new risks. This could lead to superficial compliance, adding to potential risks. More serious consequences include evasive tactics, such as creating off-the-books operations or using covert methods to avoid strict regulations. Entities might also reduce their transparency by openly sharing less information to avoid regulatory repercussions.

The timing of adopting statutory regulations is crucial. Non-statutory regions can learn from the consequences of these regulations and adapt their approach for more pragmatic implementation. Established statutory regulations might also influence non-statutory regions through processes of isomorphism, potentially leading to homogenisation or divergence (Beckert, 2010).

> Established statutory regulations might also influence non-statutory regions through processes of isomorphism, potentially leading to homogenization or divergence.

8 A Computational Analysis of Enforceability of the Data Requirement

This section examines the enforcement mechanism, especially the computational complexity entailed in enforcing the data governance requirements stipulated in the EU's AI Act (focusing mainly on Article 10 of the EU AI Act). By drawing analogies with computational NP problems, we explore the inherent challenges in meeting these requirements and the implications for compliance. The analysis highlights the evolution of the wording of the AI Act and its impact on the practical enforceability of data standards.

In computing, the enforceability of a requirement can be analogous to analysing computational complexities of nondeterministic polynomial (NP) problems. Existing technical analyses, such as those by Liza (2022), reveal the difficulties in enforcing the requirements of the 2021 draft version of the AI Act. We further explore these challenges in the context of the 2024 version of Article 10 that is currently in force.

The 2021 draft version of Article 10(2(e)) of the AI Act required datasets to be: "relevant, representative, free of errors and complete". Liza (2022) analysed the enforcement difficulties of this requirement, noting the inherent impossibilities in making sure that datasets meet strict criteria. The 2024 version of Article 10(3) presents a modified narrative: "Training, validation and testing data sets shall be relevant, sufficiently representative, and to the best extent possible, free of errors and complete in view of the intended purpose. They shall have the appropriate statistical properties, including, where applicable, as regards the persons or groups of persons in relation to whom the high-risk AI system is intended to be used. Those characteristics of the data sets may be met at the level of individual data sets or at the level of a combination thereof". Given the limited scope of this chapter, let us assess the compliance complexity with the requirement that a dataset be "sufficiently representative" by formulating as follows:

Problem: Ensure that the dataset is 'sufficiently representative' in view of the intended purpose (e.g., of the target population).

Complexity: This involves selecting a subset of the data that accurately represents the distribution of the target population, which can be seen as a variant of the Set Cover problem (Lund & Yannakakis, 1994). The Set Cover problem, especially when aiming for specific statistical properties, is known to be NP-hard.

Entities claiming to be in compliance must demonstrate they have solved an NP-hard problem, which is very difficult, if not impossible. In practice, there are approximate algorithms to solve such a problem which does not guarantee an exact or correct solution. The elaborate illustration and proof of this lie beyond the scope of this chapter.

One argument is that legal rules and principles can be ambiguous, open to multiple interpretations, or not clearly applicable to every situation (Kress, 1989, Endicott, 1996). This concept suggests that laws are not always precise or definitive, leading to uncertainty in how they should be applied or interpreted in given cases. The reasonableness (MacCormick, 1998; Bongiovanni et al.,

2009) standard is a valuable tool for mitigating the effects of legal indeterminacy.

Both approximate algorithms for NP-hard problems and reasonableness standards in law serve to manage complexity and provide practical solutions where exact answers are impractical or impossible. While approximate algorithms provide quantitative bounds on performance, reasonableness standards rely on qualitative judgments to achieve fair outcomes. This comparison highlights how both fields use pragmatic approaches to handle intractable issues, emphasising the importance of flexibility and practicality in decision-making processes. It will be interesting to see how transdisciplinary research progresses considering the new AI era to find a meaningful solution for a safe and secured future.

”The EU AI Act aims to regulate AI use in a manner that is responsible and does not stifle innovation or infringe upon human rights.

9 AI Regulation, Unintended Consequences with Merton’s Framework

The intricate implications of cross-cutting challenges (CC1, CC2, CC3), interoperability between divergent regulatory frameworks, human rights implications, and both statutory and non-statutory regulations, as well as technical enforceability challenges are becoming increasingly evident. The EU AI Act aims to regulate AI use in a manner that is responsible and does not stifle innovation or infringe upon human rights. However, Robert K. Merton’s concept of unintended consequences is particularly relevant for understanding the potential impacts of global regulatory approaches including the EU AI Act. Merton defined unintended consequences as outcomes that are not anticipated or intended by purposeful action (Merton, 1936). He identified several factors (e.g., Ignorance, Error, **Imperative of Immediate Interests**, Basic Values, Self-defeating Predictions) leading to these outcomes, which are highly relevant to the key parameters (timing, scope, method) of AI regulations:

A. **Ignorance:** A lack of comprehensive knowledge about the full range of disciplines involved in AI regulations (see section 5, CC3) can lead to significant unintended consequences. These include challenges in global cooperation, as well as divergent policy development and enforcement strategies. A transdisciplinary approach is critical to avoid gaps in understanding, conceptualising and implementation.

B. **Error:** Incorrect assumptions about the relationships between regulatory actions (such as the feasibility of technical standards outlined in section 5, CC1) and their outcomes can result in compliance inconsistencies (see section 6). Further, the technical and legal infeasibility of enforcement mechanisms (see section 8) can undermine regulatory efforts.

C. **Imperative of immediate interests:** The pursuit of short-term goals, such as gaining political popularity or asserting supremacy, can overshadow the long-term effects of regulations. This can lead to human rights violations, particularly affecting marginalised groups (see section 6). It is crucial to balance immediate interests with long-term consequences to ensure ethical governance.

D. **Basic values:** Actions influenced by fundamental values, such as autonomy and ethics (see section 5, CC2), can sometimes neglect the broader consequences. For instance, while ethical considerations are vital, they must be integrated with practical technological feasibilities to avoid unrealistic regulatory expectations and to avoid encouraging loophole pursuit.

> A lack of comprehensive knowledge about the full range of disciplines involved in AI regulations can lead to significant unintended consequences.

E. **Self-defeating predictions:** Actions taken to avoid predicted negative outcomes can paradoxically bring them about. For example, anthropomorphic views of AI might lead to regulatory measures that inadvertently result in human rights violations for migrants (see section 6) or induce undesired behavioural changes (see section 7). Recognising and mitigating these counterproductive outcomes is essential for effective regulation.

By initiating discussions based on these factors and expanding to include additional relevant considerations, we can better anticipate and address the unintended consequences of AI regulations. This proactive approach will help in crafting regulations that not only mitigate risks but also foster innovation and protect human rights.

10 Conclusion

This chapter aims to promote discussion and inspire for inclusive transdisciplinary collaboration to manage artificial intelligence (AI) technology to effectively mitigate unavoidable consequences – specifically the perverse result. The regulation of AI is a multifaceted challenge that demands a comprehensive, transdisciplinary approach to ensure systemic safety, fairness, and innovation for economic and social progress. As AI systems become ever more integrated into various aspects of society, the need for robust, interoperable and systemically safety promoting regulations becomes more pressing. Effective AI regulation calls for a deep understanding that spans formal, natural, and empirical sciences, social sciences, and engineering disciplines, assuring that the technology is developed and deployed responsibly.

> ”
> The regulation of AI is a multifaceted challenge that demands a comprehensive, transdisciplinary approach to ensure systemic safety, fairness, and innovation for economic and social progress.

The EU AI Act highlights the need for safe, secure, trustworthy and ethical AI, stressing requirements for explainability, transparency, accountability, and appropriate data governance. However, meeting these requirements is challenging with the current scientific capabilities. Satisfying such requirements is often made difficult due to the lack a lack of interdisciplinary comprehension among various domains. Addressing this gap through transdisciplinary approaches and collaboration is crucial for developing regulations that can effectively mitigate potential risks and unintended consequences. To achieve pragmatic and enforceable AI regulation, technocrats and policymakers must embrace a holistic view that incorporates diverse perspectives and expertise. By fostering inclusive collaboration, we can create a regulatory framework that not only addresses the technical aspects of AI but also considers the societal implications, ultimately leading to safer and more reliable AI systems.

In conclusion, in this chapter we have reflected on past unintended consequences of regulations and technology adoption. We have outlined the need for striking a balance in key parameters – timing, scope and methods – while regulating evolving AI technologies. We have categorised some key cross-cutting challenges and the unintended consequences that might have a long-term negative impact on our society and economy. The path to systemic AI safety regulations involve integrating sophisticated techniques across multiple disciplines, ensuring interoperability across borders, and implementing a pragmatic enforcement plan. Doing this will allow us to mitigate the potential risks of AI and unintended consequences of regulations, fostering an environment where AI technologies can thrive safely, responsibly, legally and ethically.

> ”
> The EU AI Act highlights the need for safe, secure, trustworthy, and ethical AI, stressing requirements for explainability, transparency, accountability, and appropriate data governance.

REFERENCES

- Abraham, J., & Davis, C. (2007). Interpellative sociology of pharmaceuticals: problems and challenges for innovation and regulation in the 21st century. Technology Analysis & Strategic Management, 19(3), 387-402.
- Beckert, J. (2010). Institutional isomorphism revisited: Convergence and divergence in institutional change. Sociological theory, 28(2), 150-166.
- Begishev, I. R., Latypova, E. Y., & Kirpichnikov, D. V. (2020). Artificial Intelligence as a Legal Category: Doctrinal Approach to Formulating a Definition. Actual Probs. Econ. & L., 79.
- Bongiovanni, G., Sartor, G., & Valentini, C. (Eds.). (2009). Reasonableness and law (Vol. 86). Springer Science & Business Media.
- BTO. (2024, July 4). What does the UK general election mean for AI regulation? Retrieved July 14, 2024, from https://www.bto.co.uk/blog/what-does-the-uk-general-election-mean-for-ai-regulation.aspx
- Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. Telecommunications policy, 44(6), 101937.
- Church, K., & Liberman, M. (2021). The future of computational linguistics: On beyond alchemy. Frontiers in Artificial Intelligence, 4, 625341.
- Wendehorst, C. (2020). Strict Liability for AI and other Emerging Technologies. Journal of European Tort Law, 11(2), 150-180. https://doi.org/10.1515/jetl-2020-0140
- Chataway, J., Tait, J., & Wield, D. (2006). The governance of agro-and pharmaceutical biotechnology innovation: public policy and industrial strategy. Technology Analysis & Strategic Management, 18(2), 169-185.
- Editors, TheFamousPeople.com. (2022, October 03). Robert Durst Biography. TheFamousPeople.com. Retrieved July 11, 2024, from https://www.thefamouspeople.com/profiles/robert-durst-52922.php
- Endicott, T. A. (1996). Linguistic indeterminacy. Oxford J. Legal Stud., 16, 667.
- Euronews. (2023, December 15). 'Potentially disastrous for innovation': Tech sector says EU AI Act goes too far. Euronews. Retrieved July 11, 2024, from https://www.euronews.com/next/2023/12/15/potentially-disastrous-for-innovation-tech-sector-says-eu-ai-act-goes-too-far
- Faulkner, A. (2009). Regulatory policy as innovation: Constructing rules of engagement for a technological zone of tissue engineering in the European Union. Research policy, 38(4), 637-646.
- Floridi, L. (2021). The European legislation on AI: A brief analysis of its philosophical approach. Philosophy & Technology, 34(2), 215-222.
- Janssen, A. (2022). AI and Contract Performance. In L. A. DiMatteo, C. Poncibò, & M. Cannarsa (Eds.), The Cambridge Handbook of Artificial Intelligence: Global Perspectives on Law and Ethics (pp. 59–73). chapter, Cambridge: Cambridge University Press.
- Kress, K. (1989). Legal indeterminacy. Calif. L. Rev., 77, 283.

- Liza, F. F. (2022). Challenges of Enforcing Regulations in Artificial Intelligence Act – Analyzing Quantity Requirement in Data and Data Governance.
- Lund, C., & Yannakakis, M. (1994). On the hardness of approximating minimization problems. Journal of the ACM (JACM), 41(5), 960-981.
- MacCormick, N. (1998). Reasonableness and objectivity. Notre Dame L. Rev., 74, 1575.
- Marcin, S. (2024). United States approach to artificial intelligence, EPRS: European Parliamentary Research Service. Belgium. Retrieved 13th July 2024 from https://policycommons.net/artifacts/11303354/united-states-approach-to-artificial-intelligence/12188724/. CID: 20.500.12592/ksn07nm.
- Muehlhauser, L., & Helm, L. (2012). Intelligence Explosion and Machine Ethics. Amnon Eden, James H. Moor, Jhonny H. Soraker, Eric Steinhart. Singularity Hypotheses: A scientific and Philosophical Assessment.
- Ramírez, N. M. O. (2018). Inteligencia artificial: ficción, realidad y... sueños. Real Academia de Ingeniería.
- Samoili, S., Cobo, M. L., Gómez, E., De Prato, G., Martínez-Plumed, F., & Delipetrev, B. (2020). AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence.
- Schuett, J. (2023). Defining the scope of AI regulations. Law, Innovation and Technology, 15(1), 60-82.
- Shetty, S. (2023, July 14). The EU's AI Act is barreling toward AI standards that do not exist. Lawfare. Retrieved from https://www.lawfaremedia.org/article/eus-ai-act-barreling-toward-ai-standards-do-not-exist
- Stewart, R. B. (1981). Regulation, innovation, and administrative law: conceptual framework. California Law Review, 69(5), 1256-1377.
- Toby Saunders (2023, July 9) How many continents are there in the world? It depends who you ask, BBC Science Focus. https://www.sciencefocus.com/planet-earth/how-many-continents-are-there-in-the-world
- Wendehorst, C. (2020). Strict liability for AI and other emerging technologies. Journal of European Tort Law, 11(2), 150-180.

Chapter 5

The EU Regulation of Generative AI: What Can We Learn from the Controversy Between the Italian Data Protection Authority and OpenAI

Michele Farisco

1 Introduction

ChatGPT (an AI language model designed to generate human-like responses to diverse inputs in the form of prompts) was introduced in late 2022 and its use has been steadily increasing in a variety of contexts (academia, medicine and business, among others). While seen as a promising tool, its functionalities, general availability, and ease of use raise significant concerns about how this specific AI system (and generative AI more generally) might impact our society, including our social rights and freedom. Accordingly, different stakeholders, including professionals in the fields of AI, economics, ethics, public policy, the social sciences and the humanities, and in some cases lay publics are debating whether and how to best regulate its use.

Some policymakers have already taken certain steps. For instance, the Italian Data Protection Authority (IDPA) temporarily banned the use of ChatGPT in Italy over privacy and use by minors concerns at the end of March 2023. OpenAI, the developer of ChatGPT, was accused of not having conformed with the EU General Data Protection Regulation (GDPR), that is with the EU regulation on data protection. OpenAI reacted to this action first by suspending the use of ChatGPT in Italy, then adopting relevant changes which made possible to restore the normal use of the chatbot. Yet the IDPA eventually considered those changes not sufficient to address all of the contested issues. In fact, it sent a notification of breaches of privacy law to OpenAI at the beginning of 2024 accusing it of not having properly handled a number of points, including personal data processing, the risk of ChatGPT 'hallucinating' (i.e., providing wrong information), transparency, and use by minors.

This controversy between the Italian authority and OpenAI illustrates that general-purpose AI (GPAI) models, including Large Language Models (LLMs) which inform ChatGPT, are very problematic from a regulatory point of view. Also, the EU AI Act, while not including them in the first proposal, eventually paid specific attention to GPAI models.

In this chapter, I first describe the issues that according to the IDPA arise from ChatGPT (and generative AI more broadly), including a reference to work conducted by the European Data Protection Board on the same issues. I then describe how the EU AI Act covers the issues arising from AI in general and GPAI models in particular; finally, against this background, I reflect on some ethical issues emerging from ChatGPT and on their possible impact on the implementation of AI regulation in Europe, concluding that it is desirable to complement the actual regulatory approach with a broader ethical analysis (Table 1).

Table 1: Prevailing regulatory approaches to generative AI with related illustrative issues, and summary of the proposed approach to AI ethics to complement actual regulation with illustrative issues

Generative AI		
	Approaches	**Issues**
AI regulation	GDPR-based (IDPA)	Right to confidential and fully informed use of personal data
	Principles-based (EDPB)	
	Risk-based (AI-Act)	
AI Ethics	Foundational/fundamental vs practical heuristics	• Epistemic/social inequalities and unbalances • Impact on human self-perception, decision-making, and self-determination

Source: Author's analysis

2 IDPA Concerns with ChatGPT

2.1 First Round: A Dispute Notice to OpenAI

Using an "emergency procedure", on 30 March 2023 the Italian Data Protection Authority (IDPA) decided for the immediate temporary limitation of the use of ChatGPT in Italy. At the same time, OpenAI, the developer of ChatGPT, was notified that the IDPA had initiated an inquiry into the facts of the case (IDPA, 2023). The general reason for this important action was ChatGPT being uncompliant with the EU General Data Protection Regulation (GDPR). Specifically, the IDPA claimed OpenAI had a lack of transparency about the data collected through ChatGPT and concerning the purpose and eventual use of their collection.

The IDPA raised different issues related to personal data collection, processing, and use. First, users were not sufficiently informed about their data being collected; second, there was not a sufficient legal basis to justify the collection and use of personal data for training the algorithms informing ChatGPT; third, since LLMs, the technology of ChatGPT, are prone to 'hallucinate' (Farquhar, Kossen, Kuhn, & Gal, 2024), there is a high risk of providing inaccurate information about personal details of people; fourth, despite the declared possibility to use ChatGPT reserved for people aged older than 13 years, there was no actual mechanism for age verification, meaning ChatGPT could eventually also be used by very young people, which entails several potential risks.

Specifically, the IDPA asked OpenAI to take the following actions:

- Elaborate a privacy policy and publish it on the ChatGPT website in order to illustrate the methods of processing the training data, the logic behind such processing, and the rights of the interested parties to all stakeholders, including people who do not use the service, whose data had been collected and used to train the ChatGPT algorithms;
- Include a tool in the ChatGPT website with which people may exercise the right to reject the processing of their personal data, which OpenAI obtains from third parties and uses to train the algorithms and to provide the service;
- Include a tool in the ChatGPT website with which users may ask for and obtain the correction of their personal data or, if a correction is impossible, ask for and obtain the removal of such data;
- Include a link on the ChatGPT website to give access to the privacy policy so that users can read it before registering;

- Modify the legal basis by virtue of which OpenAI declared to process users' personal data to train its algorithms, removing reference to the contract and indicating as the legal basis the consent of the user or the legitimate interest of the company (according to the assessments that the company itself will carry out);
- Include a request addressed to all users when they first access the service, after its reactivation in Italy, to overcome an age-related barrier, in order to prevent access by minor users;
- Develop a plan that involves the use of suitable mechanisms to check the users' age so as to prevent access to the service by people less than 13 years of age and also by minors (over 13 years of age) who do not have explicit consent from those who exercise parental responsibility; and
- Promote an information campaign on the main Italian mass communication channels, defining the contents with the IDPA, in order to inform people that their personal data could be collected to train the algorithms and that a specific document has been published on the ChatGPT website and a tool has been made available therein with which interested parties can obtain the removal of their personal data.

The IDPA gave OpenAI 20 days to inform it about the actions taken for addressing the points above. OpenAI initially suspended the use of ChatGPT in Italy, to then restore it after some corrective measures which were considered sufficient by the IDPA. More specifically, OpenAI took the following actions:

- It elaborated and published on the ChatGPT website an information document describing the use of training data to both users and non-users whose personal data has been processed to train the algorithms and how they have been processed, also reminding that everyone has the right to reject this type of data processing;
- It further detailed the information concerning the processing of personal data and made it accessible also in the registration form prior to actual registration for the service;
- It recognised the right of all people living in Europe, even if they are not users of ChatGPT, to reject the processing of their data aimed at training algorithms, through the use of an easily accessible online form;
- It enabled anyone asking for it to delete personal information that they consider to be incorrect;
- It specified in the privacy policy that personal data will be processed in order to train the algorithms only if relevant people have not exercised the right to object and that such processing will take place on the basis of legitimate interest;

- It added an option in the welcome page reserved for already registered Italian users in order for them to declare that they are adults or over 13 years of age, a necessary condition for accessing the service again. In the event users are between 13 and 18 years old, they must declare that they have parental consent;
- It added the request for the user's date of birth in the registration form, including a registration block for users under the age of 13 and asking for confirmation of parental consent to the use of the service where users are between 13 and 18 years of age.

2.2 Second Round: The Notification of Branches of Privacy Law to OpenAI

Ten months after the temporary ban described above, which was the result of an emergency procedure, the IDPA notified OpenAI (IDPA, 2024) of breaches of data protection law. This notification, which is the outcome of the investigation procedure initiated as part of the first dispute notice, is preliminary to possible administrative fines, to be eventually decided on the basis of OpenAI's actual collaboration. The reason for this second procedure is the confirmed evidence of the lack of compliance with the EU GDPR already contested during the first emergency procedure. The IDPA gave OpenAI 30 days to react to this decision, providing relevant counter-arguments. A final decision has still to be made.

The notification of privacy law breaches was made at the beginning of 2024, a particularly sensitive time for AI regulation in Europe since it was the moment when the AI Act had to be approved and it was eventually decided to include more stringent limitations on the use of generative AI (i.e., the technology informing ChatGPT) due to its potential impact on the general population. Specifically, the points that the IDPA had raised with OpenAI were the legal basis for the treatment of personal data, the risk of ChatGPT 'hallucinating' (i.e., providing wrong information on individuals with related inappropriate use of personal information), risks related to a lack of transparency and to use by minors.

Interestingly, the IDPA made an explicit reference to the work of the ad-hoc task force set up by the European Data Protection Board (EDPB). In fact, the IDPA will take this work into account in the final decision on the case in question. Among the outputs of the EDPB taskforce, the Report published on 23 May 2024 is especially relevant (EDPB, 2023). The main points of the report are reported below.

3 Report of the European Data Protection Board Taskforce on ChatGPT

Before introducing some ethical and legal principles and reflecting on their application to ChatGPT, the Report of the work undertaken by the ChatGPT Taskforce provides a working definition of General Pre-trained Transformer (GPT) systems as "general-purpose AI models according to the definition laid down by Article 3(63) of the EU Artificial Intelligence Act". This definition does not aim to be comprehensive or technically detailed, but to contextualise the Report within the framework of the EU AI regulation as expressed in the AI Act.

Importantly, the Report recognises that several national supervisory authorities (like the IDPA) have initiated data protection investigations pursuant to Article 58(1)(a) and (b) GDPR against OpenAI. This reference to the work conducted by local authorities is very important since one of the Report's main potential contributions is to establish the conditions for a coordinated effort among the different national bodies in order to end up with a consistent and hopefully unitary regulation. In fact, the Report states explicitly that "the European Data Protection Board [...] on 13 April 2023 decided to establish a taskforce to foster cooperation and exchange information on possible enforcement actions on the processing of personal data in the context of ChatGPT. [...] it was in particular necessary to coordinate national cases". Further, recognising that different investigations are still ongoing, the Report specifies that the considerations included are to be regarded as preliminary views.

The Report explicitly refers to the following principles:

- Accountability
- Lawfulness
- Fairness
- Transparency (and information obligations)
- Data accuracy.

In addition to these principles, the Report makes reference to the rights of the data subject. Table 2 summarises the principles above and related requirements for the data controllers (e.g., OpenAI in the case of ChatGPT).

Table 2: Principles endorsed by the European Data Protection Board's ChatGPT Taskforce and related requirements concerning the data controller's conduct

Principle	Requirement for the data controller
Accountability	Data controllers should apply the principle of data protection by design. "[...] controllers processing personal data in the context of LLMs shall take all necessary steps to ensure full compliance with the requirements of the GDPR"
Lawfulness	The processing of personal data should be compliant with Article 6(1) and when possible with additional requirements stated in Article 9(2) of the GDPR. "It is useful to distinguish the different stages of the processing of personal data. In the present context, the stages can be categorised into i) collection of training data (including the use of web scraping data or reuse of datasets), ii) pre-processing of the data (including filtering), iii) training, iv) prompts and ChatGPT output as well as v) training ChatGPT with prompts"
Fairness	"It has to be recalled that the principle of fairness pursuant to Article 5(1)(a) GDPR is an overarching principle which requires that personal data should not be processed in a way that is unjustifiably detrimental, unlawfully discriminatory, unexpected or misleading to the data subject"
Transparency	"[...] it is of particular importance to inform data subjects that the aforementioned 'Content' (the user input) may be used for training purposes"

Source: Author's analysis based on the EDPB (2023). Report of the work undertaken by the ChatGPT Taskforce. Brussels: European Data Protection Board

Accountability requires that data controllers are responsible for ensuring full compliance with the GDPR. In particular, they should apply the principle of data protection by design (i.e., including relevant measures to protect data confidentiality already in the very early phases of the system's development) both when determining the means and modality of processing and when data processing is actually executed. With reference to this point, the Report acknowledges that OpenAI updated their privacy policy on 15 December 2023, effective by 15 February 2024.

Lawfulness requires that the processing of personal data is compliant with Article 6(1) and when possible with additional requirements arising from Article 9(2) of the GDPR. Importantly, the Report outlines that it is useful to distinguish five stages of personal data processing: "i) collection of training data (including the use of web scraping data or reuse of datasets), ii) pre-processing of the data (including filtering), iii) training, iv) prompts and ChatGPT output as well as v) training ChatGPT with prompts". Specific attention is paid to "web scraping", which is the automated collection and extraction of information from publicly available sources on the Internet in order to use it for training ChatGPT. The legal basis for this procedure that OpenAI refers to is Article 6(1)(f) of the GDPR: "Processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party, except where such interests are overridden by the interests or fundamental rights and freedoms of the data subjects which require protection of personal data, in particular where the data subject is a child". Accordingly, three conditions are needed to justify web scraping: i) a legitimate interest by the data controller, ii) need to process personal data, and iii) a balancing of interests. The point being that a balance should be found between the data controller's legitimate interest and the fundamental rights and freedoms of the data subjects.

Fairness requires that the data controller should process personal data avoiding any detrimental, discriminatory, unexpected or misleading procedure. Finally, transparency means that data subjects should be informed about the possibility of their personal data being used for training purposes.

Alongside these principles, the Report refers to the following rights of data subjects:

- To access personal data,
- To be informed regarding how personal data is processed,
- To delete, rectify, or in certain conditions, transmit personal data to a third party,
- To restrict the processing of their data or to file a complaint with a Supervisory Authority.

The Report further stresses the need for data subjects to be able exercise these rights in a straightforward manner.

4 The AI Act's Regulation of Generative AI

The AI Act is a landmark of the EU's regulation of AI, becoming the first international attempt to provide a set of comprehensive, horizontal and harmonised rules (Parliament, 2024b). The AI Act was proposed by the European Commission in April 2021, with a first draft being agreed upon in December 2023 after a process that, among other things, included discussions between different EU member states, consultation with the public, and the involvement of relevant stakeholders. The AI Act was eventually adopted by the European Parliament in March 2024 and endorsed by the European Council in May 2024. The AI Act was formally signed on 13 June 2024 and was published in the EU's Official Journal on 12 July 2024 as Regulation (EU) 2024/1689 (Parliament, 2024b), to finally enter into force 20 days after its publication (Parliament, 2024a).

The AI Act is the last step in a long process of reflection and discussing how to regulate AI in Europe. Over the last 5 years, a number of documents have been published, like the EC Communication on Building Trust in Human-Centric Artificial Intelligence (Commission, 2019), the EU White Paper on Artificial Intelligence (Commission, 2020; EDPS, 2020), the EU High Level Expert Group on Artificial Intelligence Guidelines for Trustworthy AI (HLEG, 2019a), the Policy and investment recommendations for trustworthy AI (HLEG, 2019b), and the EC Communication on fostering a European approach to AI (Commission, 2021). In short, the AI Act employs a risk-based approach to AI, which distinguishes four levels of risks raised by AI and defines relevant rules. Table 3 describes the levels of risks and related prescriptions.

Table 3: Levels of risks identified by the EU AI Act and related prescriptions

Level of risk	Description	Prescription
Unacceptable	Violation of EU fundamental rights and values	Prohibition
High	Impact on health, safety or other fundamental rights	A number of mandatory requirements, including conformity assessment and post-market monitoring
Transparency	Risk of impersonation, manipulation or deception	Information and transparency obligation
Minimal	No particularly sensitive risk	No specific obligation

Source: Author's analysis based on the European Parliament (2024). REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Brussels: Official Journal of the European Journal

The identified risk levels are:

- Unacceptable risk (i.e., the AI system violates EU fundamental rights and values, and thus is prohibited);
- High-risk (i.e., the AI system impacts health, safety or other fundamental rights, and is therefore subject to several requirements like conformity assessment and post-market monitoring);
- Transparency risk, which also applies to chatbots (i.e., the AI system raises the risk of impersonation, manipulation or deception, and it thus subject to an information and transparency obligation); and
- Minimal risk (i.e., the AI system does not raise any particularly sensitive risk, and is therefore not subject to any specific regulation).

Alongside the four levels of risk described above, the AI Act pays specific attention to General Purpose AI (GPAI) models. GPAI (Triguero, Molina, Poyatos, Del Ser, & Herrera, 2024) or foundation models (Bommasani et al., 2022) are AI tools trained on a broad set of unlabelled data that are extremely flexible (i.e., they can be used for a number of different tasks with minimal fine-tuning) (Parliament, 2023a). Research on GPAI is among the sectors featuring the highest level of financial investment, estimated at USD 1.7 billion between 2020 and 2023 (Wiles, 2023). GPAI models inform Generative AI, that is a technology designed to generate different kinds of new content in response to a user's prompt (Parliament, 2023b; Zewe, 2023).

While not included in the European Commission's original proposal, GPAI models were eventually included in the final text following the proposal of the European Council (EU, 2022). Pursuant to the AI Act, GPAI models are subject to transparency requirements and, if considered to raise systemic risks, also to risk assessment and mitigation. A GPAI model is considered to raise systemic risk if it exceeds 10^25 floating-point operations per second (FLOPs).

More specifically, transparency requirements refer to the obligation to disclose the AI origin of the contents provided, which is an issue having an increasing societal impact, adding to the need for relevant regulation (Hacker, Engel, & Mauer, 2023), with some scholars arguing it is necessary to include detection mechanisms (e.g., watermarking (Parliament, 2023b)) before allowing the public release of the technology (Knott et al., 2023). Further, according to the AI Act, the providers of GPAI models are required to employ a policy in order to comply with EU copyright law, and they should also make publicly available a sufficiently detailed summary of the training data used for their models.

5 On the Need to Complement Regulation with a Broader Ethical Analysis

As summarised above, the main concerns about generative AI, GPAI models, and ChatGPT expressed by the EU AI Act, the Report of the EDPB Taskforce on ChatGPT, and by the IDPA actions against OpenAI generally arise from the risks of impacting the confidentiality of personal information and of exploiting personal information for unauthorised uses. Accordingly, the transparency risk identified by the AI Act, as well as the principles of accountability, lawfulness, fairness and transparency referred to by the EDPB Taskforce ChatGPT Report, and the IDPA procedures against OpenAI revolve around the individual right to confidential use of personal information and the user's right to be fully informed about how their data is handled and possibly used. Indeed, these are important regulatory and ethical dimensions, and it is surely a priority to assess them in response to the growing use of these technologies. Yet these issues do not exhaust the ethical relevance of generative AI generally and ChatGPT in particular, and they should be complemented with broader ethical reflections.

The considerable controversy surrounding the new frontiers of AI technologies (e.g., generative AI) is illustrated by the attention it has garnered among both professionals and the general public. For instance, several stakeholders signed an Open Letter calling for a 6-month pause in the research and training of AI technology systems like those underlying ChatGPT (i.e., GPT-4) in order to allow the strongly needed and more effective ethical assessment of it and, if not implemented, for a governmental moratorium (FLI, 2023). Nonetheless, the efficacy as well as the necessity of this call for a moratorium on research on GPT-like technologies is questionable.

A tailored ethical reflection on ChatGPT requires context. In fact, ChatGPT is only the tip of the iceberg, where the iceberg is AI technology that relies on Artificial Neural Networks (ANNs), notably Deep Neural Networks (DNN), which currently represent the state-of-the-art performance in different Machine Learning (ML) fields. This means the ethical challenges ChatGPT raises are not disconnected from the ethical issues raised by AI technology in general, even if this innovation presents them in a more obvious and attention-grabbing way. In fact, in knowledge-based and knowledge-driven societies like those we live in, the potential impact of ChatGPT, based on processing data for deriving knowledge which is relevant to the needs/interests of the user, is theoretically huge (Danaher, 2024). ChatGPT risks further emphasising already existing epistemic inequalities and unbalances, which depend importantly on unequal access to technologies, not only because they are not physically available or not economically affordable, but also since, despite the theoretical wider availability of the technologies, people do not know how best to use them so as to maximise the resulting benefit. In this way, epistemic unbalance translates into increased social inequalities, eventually impacting individual rights and freedoms (e.g., to flourish).

If the above is true, does the general availability and use of ChatGPT lead to any specific ethical challenges? Even though it is probably hard to argue that it entails unique challenges, it could be argued that the system can exacerbate existing ones (Dwivedi et al., 2023). The ethical discussion on AI, including its social and political implications, has seen significant development in recent years (Coeckelbergh, 2020; Dignum, 2019; Stahl, 2021) as concerns the increased technical capacities of AI systems and expanded scope of their application. Focusing on the specific case of brain-inspired AI, I recently introduced a method for the ethics of AI, which is proposed to distinguish two main kinds of ethical issues: foundational/fundamental (i.e., concerning the impact AI has on our fundamental moral notions, like responsibility, freedom etc.) and practical (i.e., concerning the impact AI has on a number of issues in different contexts, including how to safeguard individual rights and maximise the societal benefit deriving from AI)(Farisco et al., 2024). More specifically, we may identify at least two main categories of foundational ethical issues: those related to goals (e.g., what is AI developed for?) and those related to concepts (e.g., what are the underlying categories, including morally relevant notions, used in AI research?). The practical issues can be organised in terms of the following main levels:

- *Operational*, related to how AI works;
- *Instrumental*, related to how people use AI;
- *Relational*, related to how people see AI and to the resulting psychological and metaphysical human-AI relationship; and
- *Societal*, related to the social and economic costs and consequences of the development and use of AI.

Strategy used in the ethical reflection on ChatGPT (e.g., the risks and benefits of using it in different sectors) is to compare its functionalities with human cognitive capabilities and to stress emerging differences, eventually taking them as evidence of ChatGPT's shortcomings in various contexts. For instance, it has been stressed that ChatGPT is still brittle and fragile, as shown by the gross mistakes it makes, and that it lacks the sophistication and flexibility of human intelligence. Accordingly, the comparison between ChatGPT and human intelligence has been crystallised in the dichotomy between the iterative data-driven processing characterising ChatGPT and the ´creativity´ characterising human intelligence. Part of this comparative analysis revolves around ChatGPT's lack of specific features, such as understanding and thinking, and the need for grounding (i.e., to formalise and operationalise the connection between symbols and their meaning) (Lake & Murphy, 2023). However, to focus on this comparison obscures the fact that the most challenging ethical risk arising from ChatGPT is not whether and how much it differs from human intelligence, but the impact it may have on human self-perception, decision-making, and self-determination regardless of the fact that it lacks fundamental human features. The reality is that ChatGPT has not been developed to possess actual human characteristics, but to simulate or counterfeit them (Dennett, 2019).

In the light of the above, what are the ethical issues that appear to be exacerbated by ChatGPT? The technical limitations of ChatGPT, both intrinsic to its architecture and depending on the actual technological stage, become ethically salient when combined with the human inclination to anthropomorphise AI (Salles, Evers, & Farisco, 2020). The human predisposition for over-attribution is intensified by the fact that GPT is designed to generate human-like answers to questions and prompts. Although technology that pretends to be human is not new in human history, contemporary AI systems, and ChatGPT in particular, are quite advanced in this respect, and their ease of use makes them potentially more pervasive and invasive than other technologies. Moreover, the spectrum of activities ChatGPT may be used for is much wider than traditional AI, and promises to further expand, often in ways not completely transparent to its users. This adds to the risk for misplaced trust and increased epistemic and axiological delegation, i.e., relying on ChatGPT to gain knowledge and make choices, including moral ones (Krugel, Ostermaier, & Uhl, 2023). While it may be questioned whether relying on AI to gain knowledge and make moral choices is better or worse than relying on other humans, it is a fact that the growing role of AI in replacing human actors raises the possibility of important changes in our social relationships. The case of attributing the capacity for conscious experience to LLM like ChatGPT is telling of the impact the technology has on fundamental notions, including ethically relevant concepts (Chalmers, 2023; De Cosmo, 2022; Shanahan, 2024).

Specifically, ChatGPT functionalities and the human tendency to anthropomorphise AI can lead to attributing three ethically salient features to ChatGPT: competence, judgment and agency. This possible attribution is not trivial, however. It might considerably impact three dimensions of human identity which play a significant social role: (self)-experience (i.e., (self-) knowledge); choice (i.e., decision-making); action (i.e., behaviour). Given this potential impact, in practice AI systems may be seen not just as socio-political factors (i.e., a passive variable to take into consideration) but as new socio-political actors, that is perceived by humans as purposeful agents with causal power within the symbolically and emotionally charged relationships with them, contributing to the creation of our societal and cultural worlds, eventually falsely simulating reality (i.e., giving the wrong perception that the information it provides is fully reliable, while AI is prone to hallucinating or producing 'bullshit' since it is indifferent to the truth of its outputs (Hicks, Humphries, & Slater, 2024)).

How to address the fundamental and practical ethical issues pertaining to ChatGPT? There is probably there is no single correct answer to this question. As noted above, the Open Letter calls for a pause in the specific line of research that led to GPT-4, the technology behind the current ChatGPT. This option could allow for a more careful and in-depth ethical reflection on a technology which seems far too fast for ethics to stay updated. However, not only is the feasibility of the proposal questionable (insofar as there are plenty of factors pushing for a restless ride towards even more advanced GPT-4-like technology, including economic and (geo) political interests); it is unclear whether that this is the best way forward (Ienca, 2023).

Rather than stopping the research, a different option is to change the process of AI research and development itself, paying more attention to and optimising possible regulatory approaches about the motivations, priorities and goals of investors and developers. This might be pursued, for instance, through dedicated efforts, endorsed and promoted by public authorities, at public consultation and engagement, as exemplified by the EU AI Act writing process, and could result in a different trajectory of AI development coordinated with socio-political stakeholders, potentially with very effective results.

> AI systems may be seen as new socio-political actors, perceived by humans as purposeful agents with causal power within the symbolically and emotionally charged relationships with them.

Among the conditions for adding to the chances of making the strategies above successful, clarifying the meaning of fundamental moral notions and elaborating clear operational recommendations are vital. For instance, the notion of responsibility should be carefully assessed. In principle, it is possible to understand responsibility as either based on collective processes and practices or in more individualistic terms.

The collective understanding distributes responsibility among diverse stakeholders like researchers, developers, innovators, funders, policymakers and users. It calls for engagement with diverse audiences about general goals, directions, the assessment of the relevant technology, and for embedding ethical and legal reflections in AI research and development from the very beginning, with the final goal of defining a framework for making responsibility an integral part of the process. The ethics-by-design and responsibility-by-design paradigms (Forum, 2020; Stahl et al., 2021), along with the work of the Ethics and Society team in the EU Human Brain Project (Aicardi et al., 2021) take this collective understanding of responsibility as a starting point. This is in principle a promising strategy, yet concrete criteria for its implementation, especially within the private sector, and in particular for avoiding the risk of ethics washing (Schultz, Conti, & Seele, 2024), remains to be further developed.

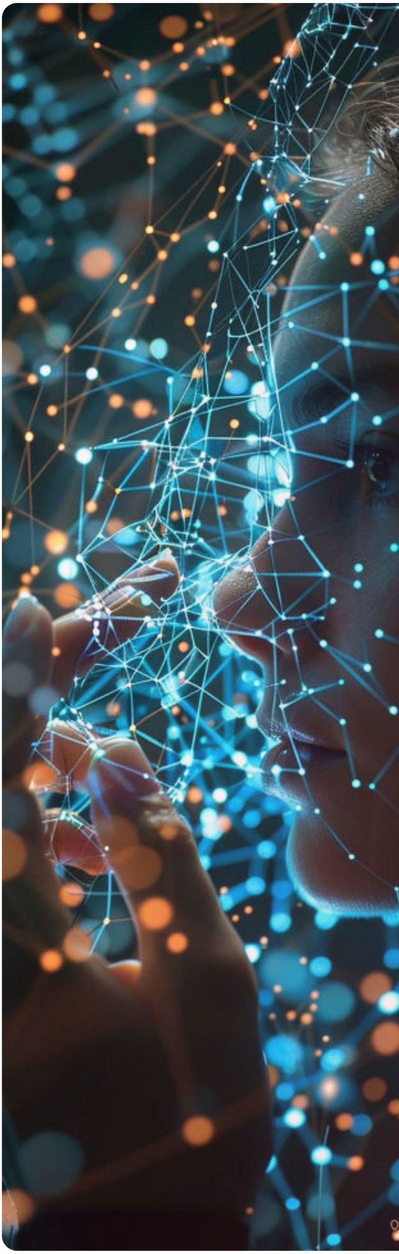
> Implementing responsible design and use of AI would require raising researchers, engineers, and developers' awareness about their social responsibility and publicly calling for it, with the ultimate goal of clarifying who is responsible for what.

From a more individualistic perspective, responsibility for the research and its direction is attributed to a particular set of people, in this case, AI researchers, engineers and developers. In this view, implementing responsible design and use of AI would require raising the awareness of these people about their social responsibility, and publicly calling for it, with the ultimate goal of clarifying who is responsible for what. Still, this strategy might be insufficient if not complemented with concrete and operationalisable recommendations in order to eventually translate awareness into effective action.

Another possibility is to strive for a more agile and open-ended AI regulation. It is true that significant illustrations in this direction are already available, especially from Europe and the USA, albeit there remains a distance between regulatory bodies and AI actors that must be reduced and hopefully filled. While other strategies might also be conceivable, probably the most effective choice is to combine different approaches. Two preconditions for a successful ethical assessment of ChatGPT are: a balanced view of the system, considering both what is risky and valuable to individuals and society at large; and awareness of the need for a broad ethical analysis, which includes but is not limited to personal data confidentiality and transparency. Importantly, in order to be as effective as possible, this ethical analysis should be translated into regulatory strategies and tools. This makes it necessary to expand the ethical scope of regulation on one hand and further operationalise ethical reflection in actual operational standards on the other.

6 Conclusion

Actual regulation of GPAI models tends to revolve around two fundamental rights of users: the right to confidentiality, and the right to be fully informed about possible uses of their data. The prevailing focus of actual regulation is thus about privacy and data protection, including the need for transparency regarding data use, and related risks assessment and mitigation. This is illustrated by the IDPA controversy with OpenAI and confirmed on the European level by the EDPB Taskforce on ChatGPT Report and the EU AI Act. Indeed, while these are important dimensions of the regulatory



relevance of ChatGPT (and generative AI more broadly), they do not exhaust its ethical significance. In fact, several additional issues arise on the ethical level, both foundational/fundamental and practical issues, which are important to acknowledge and eventually translate into a broader regulatory framework. Among them, two fundamental ethical issues may have a direct impact on classical liberal values: the new forms of epistemic and social inequalities and unbalances that may emerge from generative AI, and its possible impact on human self-perception, decision-making and self-determination.

Therefore, building on the above, I propose the following policy recommendations:

- To start from a balanced view of both GPAI models' potential benefits and risks;
- To promote the necessary conceptual clarity so that the terms used by researchers and developers, particularly those playing a key role and prone to a misleading interpretation by the general public or those that tend to raise hyped expectations (e.g., creativity, understanding etc.) are clearly defined;
- To promote the cautious use of language by researchers and developers while communicating findings so that it is adapted to relevant publics;
- To promote the creation of communication strategies sensitive to the needs and possibilities, including epistemic limitations, of different publics, which means they must be multifaceted and multi-channel;
- To further enhance efforts towards public consultation and engagement activities, including different relevant stakeholders (e.g., investors, developers, researchers, industries, members of the public), for moving towards the coordinated development of AI, in an attempt to maximise the resulting social benefit;
- To complement the risk-based approach as endorsed in the EU AI Act with an agile and open-ended regulation, thinking about tools and strategies to adapt the general principles to the particular situations at issue;
- To recognise the need for a more encompassing ethical analysis, which should include but not be limited to personal data confidentiality and transparency;
- To recognise in particular that GPAI models may have a direct impact on society, including the generation of new forms of epistemic and social inequalities and unbalances; and
- To also pay attention to how GPAI models may impact fundamental human abilities which possess a prominent democratic and liberal value, like self-perception, decision-making and self-determination.

REFERENCES

- Aicardi, C., Bitsch, L., Datta Burton, S., Evers, K., Farisco, M., Mahfoud, T., . . . Ulnicane, I. (2021). Opinion on Trust and Transparency in Artificial Intelligence – Ethics&Society, The Human Brain Project. Zenodo. Retrieved from https://doi.org/10.5281/zenodo.4588648
- Bommasani, R., & al. (2022). On the Opportunities and Risks of Foundation Models. Retrieved from arXiv:2108.07258
- Chalmers, D. (2023). Could a Large Language Model be Conscious? Retrieved from arXiv:2303.07103 website:
- Coeckelbergh, M. (2020). AI ethics. Cambridge, MA: The MIT Press.
- Commission, E. (2019). Building Trust in Human-Centric Artificial Intelligence. Brussles: European Commission.
- Commission, E. (2020). White Paper on Artificial Intelligence: a European approach to excellence and trust. Brussles: European Commission.
- Commission, E. (2021). Communication on Fostering a European approach to Artificial Intelligence. Brussels: European Commission.
- Danaher, J. (2024). Generative AI and the future of equality norms. Cognition, 251, 105906. doi:https://doi.org/10.1016/j.cognition.2024.105906
- De Cosmo, L. (2022). Google Engineer Claims AI Chatbot Is Sentient: Why That Matters. Retrieved from https://www.scientificamerican.com/article/google-engineer-claims-ai-chatbot-is-sentient-why-that-matters/
- Dennett, D. C. (2019). What can we do? We don't need artificial conscious agents. We need intelligent tools. In J. Brockman (Ed.), Possible Minds: Twenty-Five ways of Looking at AI (pp. 41-53). New York: Imprint of Penguin Publishing Group.
- Dignum, V. (2019). Responsible Artificial Intelligence: How to Develop and Use AI in a Responsible Way. Verlag: Springer.
- Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., . . . Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. doi:https://doi.org/10.1016/j.ijinfomgt.2023.102642
- EDPB. (2023). Report of the work undertaken by the ChatGPT Taskforce. Brussels: European Data Protection Board.
- EDPS. (2020). Opinion on the European Commission's White Paper on Artificial Intelligence – A European approach to excellence and trust (Opinion 4/2020) (Opinion No. 4/2020).
- EU, C. o. t. (2022). Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights [Press release]. Retrieved from https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/

- Farisco, M., Baldassarre, G., Cartoni, E., Leach, A., Petrovici, M. A., Rosemann, A., . . . van Albada, S. J. (2024). A method for the ethical analysis of brain-inspired AI. Artificial Intelligence Review, 57(6), 133. doi:10.1007/s10462-024-10769-4
- Farquhar, S., Kossen, J., Kuhn, L., & Gal, Y. (2024). Detecting hallucinations in large language models using semantic entropy. Nature, 630(8017), 625-630. doi:10.1038/s41586-024-07421-0
- FLI. (2023). Pause Giant AI Experiments: An Open Letter. Retrieved from https://futureoflife.org/open-letter/pause-giant-ai-experiments/
- Forum, W. E. (2020). Ethics by Design: An organizational approach to responsible use of technology. Retrieved from Cologny/Geneva:
- Hacker, P., Engel, A., & Mauer, M. (2023). Regulating ChatGPT and other Large Generative AI Models2023 ACM Conference on Fairness, Accountability, and Transparency (FAccT '23), June 12-15, 2023, Chicago, IL, USA. New York, NY, USA: ACM.
- Hicks, M. T., Humphries, J., & Slater, J. (2024). ChatGPT is bullshit. Ethics and Information Technology, 26(2), 38. doi:10.1007/s10676-024-09775-5
- HLEG. (2019a). Ethics Guidelines for Trustworthy AI. Brussels: European Commission.
- HLEG. (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence. Brussles: European Commission.
- IDPA. (2023). Artificial intelligence: stop to ChatGPT by the Italian SA. Retrieved from https://www.garanteprivacy.it/web/guest/home/docweb/-/docweb-display/docweb/9870847#english
- IDPA. (2024). ChatGPT: Italian DPA notifies breaches of privacy law to OpenAI. Retrieved from https://www.garanteprivacy.it/home/docweb/-/docweb-display/docweb/9978020
- Ienca, M. (2023). Don't pause giant AI for the wrong reasons. Nature Machine Intelligence. doi:10.1038/s42256-023-00649-x
- Knott, A., Pedreschi, D., Chatila, R., Chakraborti, T., Leavy, S., Baeza-Yates, R., . . . Bengio, Y. (2023). Generative AI models should include detection mechanisms as a condition for public release. Ethics and Information Technology, 25(4), 55. doi:10.1007/s10676-023-09728-4
- Krugel, S., Ostermaier, A., & Uhl, M. (2023). The moral authority of ChatGPT. Retrieved from arXiv:2301.07098v1 [cs.CY]
- Lake, B. M., & Murphy, G. L. (2023). Word meaning in minds and machines. Psychol Rev, 130(2), 401-431. doi:10.1037/rev0000297
- Parliament, E. (2023a). General-purpose artificial intelligence. Brussels: European Parliament.
- Parliament, E. (2023b). Generative AI and watermarking. Brussels: European Parliament.
- Parliament, E. (2024a). Artificial intelligence act In "A Europe Fit for the Digital Age". Retrieved from https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-regulation-on-artificial-intelligence
- Parliament, E. (2024b). REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Brussels: Official Journal of the European Journal Retrieved from http://data.europa.eu/eli/reg/2024/1689/oj
- Salles, A., Evers, K., & Farisco, M. (2020). Anthropomorphism in AI. AJOB Neurosci, 11(2), 88-95. doi:10.1080/21507740.2020.1740350
- Schultz, M. D., Conti, L. G., & Seele, P. (2024). Digital ethicswashing: a systematic review and a process-perception-outcome framework. AI and Ethics. doi:10.1007/s43681-024-00430-9
- Shanahan. (2024). Talking about Large Language Models. Commun. ACM, 67(2), 68–79. doi:10.1145/3624724

- Stahl, B. C. (2021). Artificial intelligence for a better future : an ecosystem perspective on the ethics of AI and emerging digital technologies SpringerBriefs in Research and Innovation Governance, (pp. 1 online resource). Retrieved from Directory of Open Access Books https://directory.doabooks.org/handle/20.500.12854/67925
- Springer Nature Open Access eBooks https://link.springer.com/book/10.1007/978-3-030-69978-9 doi:10.1007/978-3-030-69978-9
- Stahl, B. C., Akintoye, S., Bitsch, L., Bringedal, B., Eke, D., Farisco, M., . . . Ulnicane, I. (2021). From Responsible Research and Innovation to responsibility by design. Journal of Responsible Innovation, 1-24. doi:10.1080/23299460.2021.1955613
- Triguero, I., Molina, D., Poyatos, J., Del Ser, J., & Herrera, F. (2024). General Purpose Artificial Intelligence Systems (GPAIS): Properties, definition, taxonomy, societal implications and responsible governance. Information Fusion, 103, 102135. doi:https://doi.org/10.1016/j.inffus.2023.102135
- Wiles, J. (2023). Beyond ChatGPT: The Future of Generative AI for Enterprises. Retrieved from https://www.gartner.com/en/articles/beyond-chatgpt-the-future-of-generative-ai-for-enterprises
- Zewe, A. (2023). Explained: Generative AI. Retrieved from https://news.mit.edu/2023/explained-generative-ai-1109

Chapter 6

Power Imbalances and the AI Act's National Supervisory Authorities: Bulgaria as a Case Study of the Challenges for the EU's Periphery

Fabio Ashtar Telarico

1 Introduction

The rapid spread of generative artificial intelligence (AI) amongst businesses and consumers alike has presented politicians and other decision-makers around the world with the issue of how to regulate such tools – if at all. As the dust is lifting since the initial, frantic rush to take a position "where being the quickest to say 'we're doing anything' counts" (Hern, 2024), two different vision are emerging. One has its champions on the other side of the Channel and the Atlantic, the other is embodied in the European Union (EU). Both clearly have their own drawbacks and limitations. Yet only one of the two strives to account for power imbalances between regulators and an AI "oligopoly of vertically integrated 'foundries' and monolithic players" (Marcelli et al., 2024, p. 2; Sitaraman & Narechania, 2023). Unfortunately, it is not the EU that has got it right.

The approach that, for lack of a better term, can be defined as the Anglo-Saxon way of dealing with AI regulation is ambivalent. On the outside, it proudly wears the mantle of laissez faire and fills its mouth with the neoliberal rhetoric about innovation and competitiveness. However, it conceals a strongly dirigiste and even directly interventionist role for the state. With the White Paper "A pro-innovation approach to AI regulation", authorities in the UK stated the goal of AI regulation is to create and environment in which "innovators can thrive" (Donelan, 2023). And it is not like the UK has any gap to close: the country is already the third-largest AI research hub in the world (Hankins et al., 2023). Meanwhile, the United States has close to no real, federal- or state-wide, AI regulation with most attempts being stonewalled "to promote competition and innovation" (Plotinsky & Cinelli, 2024). But behind this veil of regulation, one can sense a much more hands-on approach. In fact, these two leading global players in AI have adopted a "whole of government" (EO 14110, 2023) approach (also called "multiple sector, multiple regulator" in the UK; see Donelan, 2024). In practice, this means putting in place the infrastructure for effective, centralised coordination between existing agencies within the existing regulatory architecture.

The EU has scrapped its usual way of regulating the common market, opting instead for the maximum total harmonisation of AI regulation (Veale & Borgesius, 2021), with very few exceptions (e.g., AI Act, 2024, art. 5 para. 4 on remote biometric identification systems). By so doing, the EU's stated goal was to muster the trade-off between regulatory scrutiny and the entrepreneurial spirit of tech companies. It thus put forth concerns about safety (the 'human-centric' side of AI regulation) and tried to balance them against the perceived need to leave enough room for private investment and innovation. Still, with its AI Act, the EU has made a few faux pas that can produce market distortions such as imposing differing limitations on public and private entities, which disadvantages the former (Georgieva et al., 2022), or drawing a not-so-clear distinction between high- and low-risk uses of AI (Veale & Borgesius, 2021). Further, national regulators can neither impose additional restrictions nor relax those already imposed on the EU level (AI Act, 2024, art. 1 para. 4). Hence, the price of the AI Act's Faustian bargain between innovation and safety has been to sacrifice the national peculiarities and specific needs of individual member states (MSs), especially the poorest and least able to withstand global competition (Telarico, 2024, p. 161). Yet, the drafters of the AI Act still found it sensible to pave the way for national market surveillance authorities (MSAs), designated by each member state, to supervise the application and implementation of these common rules (AI Act, 2024, arts. 70–74). While the norms outlining MSAs' structure, duties and competences should assure they have the necessary resources and authority to enforce the AI Act effectively, many bread-and-butter issues are not mentioned. In a nutshell, the EU's approach is at its weakest exactly where the Anglo-Saxon one is at its best: effective market oversight.

These misgivings generally stem from a legalistic approach to market oversight that completely neglects the main lesson of political economy: that market power matters. The entire legislation on MSAs hinges on the quiet assumption that AI companies will not try to deceive – or, even worse – capture these national regulators. However, this naivety has resulted in a series of colossal loopholes that tech giants like Google and Microsoft can easily exploit to weaken scrutiny on the EU's periphery, where the weaknesses of MSAs are most evident. As this policy paper highlights, the deficiencies in the European approach to AI regulation are not unresolvable. And, fortunately, the experience of other democratic countries leading the AI race provides an interesting blueprint. Still, how and to what extent a whole-of-government approach can be adapted to the needs of peripheral EU countries remains unclear. This means it is useful to look at possible solutions from the perspective of a country paradigmatic of the untapped potential of the EU's periphery in the high-tech sphere: Bulgaria.

Against this background, this paper recommends that EU policymakers revise the role of MSAs by accounting for the power imbalance between overseen companies and overseeing states. Particular attention should be paid to countries like Bulgaria that are attracting considerable scrutiny in the high-tech world, but whose institutional track record proves their inability to withstand the pressure of big tech. Concretely, following the examples of the USA and the UK, it makes six recommendations in three policy areas:

Staff's honesty and skills as a counterweight to lobbyists:

- Creating an EU fund for upskilling national regulators allocated based on actual needs to pay for tech and language training as well as staff mobility and internationalisation;
- Setup rules to incentivise whistle-blowers and to put courts of law in charge of appeals to provide people with a vested interest in effective market surveillance with the power to effect change and reduce industry lobbying.

Streamlining the ecosystem around MSAs:

- Upgrading the European Artificial Intelligence Office to an agency coordinating and overseeing MSAs to guarantee consistent enforcement, prevent regulatory capture, and spread best practices. The AI Office should become the surveillance authority for EU institutions instead of the European Data Protection Supervisor;
- Abolishing the Board that would function as a venue for lobbyists to channel their influence and transfer its advisory functions to the newly empowered AI Office.

To give these improved MSA real powers:

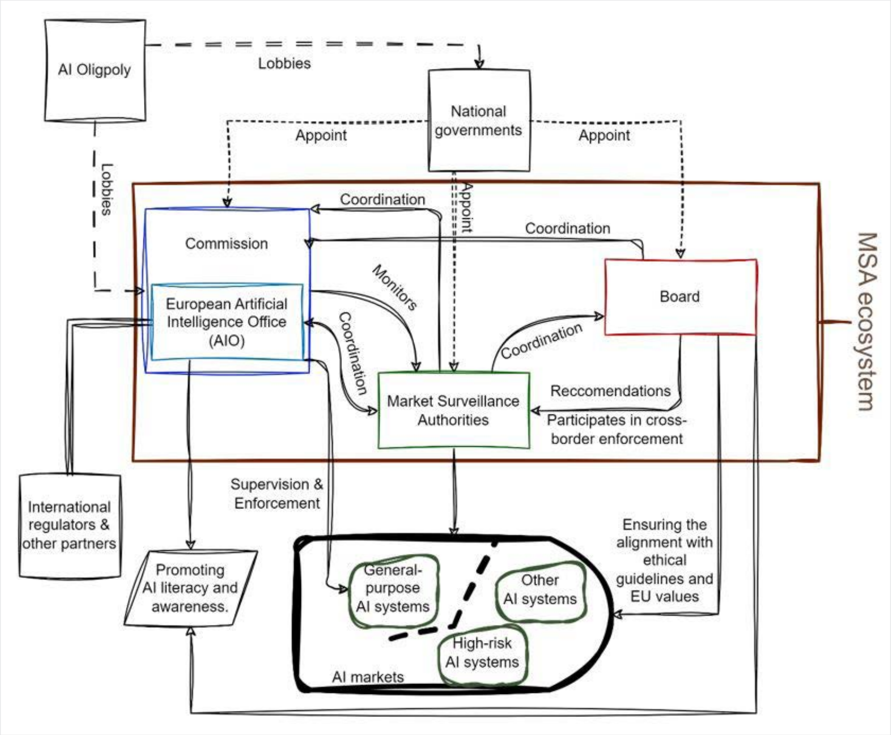
- Granting the MSAs some policy space to solve the tension between a maximally and totally harmonised regulation and its national enforcement;
- An MSA-based intergovernmental framework for public investment and acquisitions in AI.

These recommendations (detailed in Section 4) are based on a critical analysis of passages of the Al Act that establish the framework for MSAs and possible interactions with other EU-level normative acts (Section 2). Moreover, they take stake of a peripheral perspective on Al policy via an analysis of the institutional track record of supposedly independent market-overseeing authorities in Bulgaria (in Section 3).

2 MSAs in the AI Act: Disarming David as Goliath Approaches

An attentive, critical reading of the AI Act reveals that it creates a complex MSA ecosystem (Figure 1) which constrains market overseers, especially in smaller and less affluent countries. These structural weaknesses are concerning as MSAs are expected to play a crucial role in safeguarding civil liberties against the potentially overreaching powers of the AI oligopoly. Although the regulation allows MSs to either establish new agencies or empower existing ones to act as MSAs, it severely limits their autonomy by pushing them towards selecting certain pre-existing institutions, often regardless of their efficiency. This is particularly problematic since it financially burdens these nations by forcing them to comply with extensive requirements, including maintaining a highly specialised workforce and implementing robust cybersecurity measures. The Act's intricate framework also disperses coordination responsibilities across multiple bodies, reducing the effectiveness of MSAs. Further, the influence of lobbyists and overarching authority of the Commission undermine the MSAs' independence and effectiveness. In summary, underfunded and constrained as they are, the MSAs are a disarmed David that can protect neither citizens' freedoms nor the EU's founding liberal values against the AI-oligopoly Goliath.

Figure 1: The complex MSA ecosystem built by the AI Act


AI Oligpoly
Lobbies
National governments
Appoint
Appoint
Lobbies
Coordination
Appoint
Coordination
Commission
Monitors
Board
European Artificial Intelligence Office (AIO)
Coordination
Coordination
MSA ecosystem
Market Surveillance Authorities
Reccomendations
Participates in cross-border enforcement
Supervision & Enforcement
International regulators & other partners
Ensuring the alignment with ethical guidelines and EU values
Promoting AI literacy and awareness.
General-purpose AI systems
Other AI systems
High-risk AI systems
AI markets

Source: Author's elaboration based on the AI Act (2024).

2.1 (Not So) Free to Choose Your MSA/s

At first sight, the EU seems to leave the MSs with much needed flexibility in setting up and operating their MSAs (AI Act, 2024, arts. 70 and 74). Yet, in fact the AI Act virtually strongarms peripheral countries into picking specific institutions as MSAs regardless of their efficiency.

According to the letter of the regulation, national authorities can both establish new agencies or empower existing ones to take on the role of an AI-market surveillance authority. The basic precondition is that the MSA can operate independently, impartially and without bias while implementing the regulation. However, the freedom of choice and possibility of having several bureaucracies acting as counterbalances to each other is seriously limited. In fact, even if there are several MSAs, one must act as a 'point of contact' with

EU institutions. There accordingly must be a primus inter pares amongst the MSAs. Moreover, the EU seems to be pushing countries to choose particular institutions for this function. For instance, AI systems with application to the financial sectors are, by default, under the supervising authority of the body overseeing credit institutions pursuant to other EU regulations. Legally, the MSs can decide to designate another authority for these tasks, but there is a clear indication from Brussels that the EU would prefer to have the ECB ultimately in charge of overseeing financial AI. Further, the AI Act establishes the European Data Protection Supervisor to act as the MSA for Union institutions, bodies, offices and agencies. Countries are hence nudged towards selecting as the a 'general' MSA (or, at least, a contact point) their data protection agency.

As if these norms were not sufficient, the regulation adds an economic constraint on poorer countries' ability to proceed contrary to these veiled recommendations. Namely, like with other cases in the past, the cost of running the MSAs will weigh on national budgets even though the requirements are set in EU law. And this includes ensuring the agencies have a sufficient number of staffers with expertise in AI technologies, computing, personal data protection, cybersecurity, and other relevant areas. Not to mention that the MSs will also have to pay individually to satisfy the mandate to implement appropriate cybersecurity measures to protect their MSAs against potential threats. Such requirements must also be updated annually to reflect the evolving nature of AI and its associated risks: which means they will be only increasing. All things considered, the AI Act makes it virtually impossible for poorer and smaller countries to create a new agency for the role of MSA. By so doing, it may force them to select already overburdened, insufficiently qualified and generally inefficient agencies to supervise this key sector.

2.2 No Clear Way for the MSAs to Coordinate

While trying to force cooperation by making the choice of MSAs uniform, the AI Act assigns responsibilities for the coordination of MSAs to several actors, leaving no single forum for cooperation.

In fact, the regulation portrays the Commission as holding a pivotal role in facilitating the exchange of experience between national competent authorities (AI Act, 2024, art. 70 para. 7), yet it tasks the MSs with facilitating the coordination of MSAs and other relevant national authorities or bodies that supervise the application of Union harmonisation legislation (AI Act, 2024, art. 74 para. 10). In addition, it assigns the role of a permanent locus of dialogues to a standing sub-committee of the Board that serves as a platform for continuous interaction and collaboration, ensuring that MSAs can effectively share information, strategies and best practices (AI Act, 2024, art. 65 para. 6). It does so while determining that MSAs can cooperate with each other together with the Commission (AI Act, 2024, art. 74 para. 11) as well as through the Board or notwithstanding it (AI Act, 2024, art. 66 para. 1, let. a).

2.3 Sandboxing and Real-World Testing: A Trojan Horse

Although MSAs' role in sandboxing and real-world testing of a high-risk AI system is strongly emphasised, their real powers are limited (arts. 57.7, 60.4, 60.6–8, 76.2–5). Technically, MSAs can conduct unannounced (remote or on-site) inspections. They must also be notified of serious incidents and subsequent actions to mitigate or resolve the root problems. If one such notification is received (or irregularities are found), the MSAs can take enforcement actions, including suspending or terminating tests or requiring modifications to the test plan. Still, any of these decisions has to be well-motivated and subject to appeal. MSAs must additionally communicate significant decisions to other relevant MSAs, exposing them to further legal objections.

After passing the test, MSAs are encouraged to consider sandbox activity reports to expedite conformity assessments. Yet the ability to ensure that the real-world test actually provides information about the risks the system poses is limited. Crucially, testing plans are tacitly approved 30 days after being submitted (unless national law does not foresee tacit approval). Moreover, providers can also extend testing periods with a simple reasoned request, giving them time to fix possibly unreported serious incidents or any other non-conformance issue while avoiding enforcement.

2.4 Few Checks on Law Enforcement's Deployment of Non Conforming Systems

The emergency authorisations of a not-(yet-)conforming high-risk AI system is a serious concern for individual freedoms, especially given that MSAs can easily be sidestepped in several sensitive applications (AI Act, 2024, art. 46, paras 1–2, 6, Nº 130). Theoretically, it is the MSA that exercises the relatively indiscriminate power of issuing emergency authorisation in a wide range of circumstances. It is namely enough for the MSA to claim "exceptional reasons of public security or the protection of life and health of persons, environmental protection or the protection of key industrial and infrastructural assets" to make a high-risk AI system operative.

Nonetheless, at least MSAs' powers are limited and subject to judicial review. What is more concerning is that law enforcement or civil protection authorities can deploy high-risk AI systems without any authorisation from the MSA if they can argue there is a "substantial, and imminent threat to public security or the physical safety of individuals." And this use is only subject to review by the MSA "without undue delay" – essentially post-deployment. Given that the MSA is bound to respect confidentiality (Art. 78), even repeated violations of civil rights might not lead to criminal accountability and the necessary public scrutiny.

2.5 Taking the Slingshot from David's Hands: The Commission & the "Board"

Despite the apparently large remit of their powers, MSAs: (a) must service other agencies; (b) have their decisions reviewed by the Commission; and (c) are exposed to lobbyists.

As to the first point, the MSAs' remit can be eroded by national authorities' reasoned requests to participate in ad-hoc testing of high-risk AI systems suspected of violating Union law (art. 77 paras 1, 3).

Second, it is important to note that both the sanctions and the remedies MSAs can force upon the providers of high-risk AI systems in case of serious incidents during sandboxing and emergency authorisations are subject to opposition from the Commission, other MSAs and the MSs (AI Act, 2024, art. 46 para. 3). Further, the Commission is the final arbiter for all national decisions and acts as judge and jury, being able to raise objections against an MSA's decision and judge on its merits (art. 81 para. 1).

Finally, the undue influence of lobby groups is channelled to the MSAs through the "Board", a collegial body of industry representative appointed by the MSs (AI Act, 2024, № 149). In other words, special interest groups aligned with national champions will be issuing opinions, recommendations and guidance on the implementation of AI regulations. Crucially, the Board's advice extends to enforcement matters, technical specifications and existing standards, as well as recommendations concerning the MSAs' capabilities. Moreover, it provides 'inputs' that the MSAs must consider in their guidance to small and medium-sized enterprises and start-ups (AI Act, 2024, art. 70 para. 8) and must be involved in the case of cross-border investigations (AI Act, 2024, art. 66 para. 3, let. c).

2.6 A (Toothless) Defender of Civil Rights

Weakening the MSAs may leave European citizens with a disarmed David to defend their civil liberties against the Goliath of the AI oligopoly. Overall, the weak position in which the AI Act puts MSAs may seem a minor concern. But the law also tasks the MSAs to act as the prime defender of civil rights against

AI. First of all, MSAs can autonomously evaluate high-risk AI systems' impact on fundamental rights and require transparency from providers regarding their risk assessments. They can also restrict or prohibit the market availability of compliant AI systems that infringe on fundamental rights until measures are taken to mitigate the relevant risks. From a geopolitical perspective, MSAs are vital for overseeing the activities of high-risk AI system providers from (potentially hostile) third countries and should answer complaints and risk notifications from importers and deployers. Further, they gather a large amount of data that can reveal abuses in the deployment of AI. Notably, MSAs can request immediate access to documentation proving the conformity of high-risk AI systems from providers in law enforcement, immigration control, and asylum adjudication. Further, they have private access (along with the Commission) to a secure non-public section of the EU database on high-risk AI deployers, which includes actors in law enforcement, migration, asylum, and border control management along with exclusive access to the section concerning critical national infrastructures (art. 78 para. 3). MSAs also receive notifications about the use of real-time biometric data for law enforcement purposes and report annually on the topic to the Commission.

3 The View from Bulgaria on the Political Economy of MSAs

The economic literature is rife with case studies, theories, and other explanations for the failures of 'independent' agencies tasked with market surveillance. If anything, the study of overseen corporations' capture of overseeing agencies won George Stigler a Nobel Prize (Nobel Prize Outreach, 1982). All of these textbook flaws are present in the current design of the MSA ecosystem and compounded in countries with weak institutions. Indeed, in spite of renewed emphasis on market surveillance in the EU, the inherent structural faults of executive agencies remain unsolved and only exacerbate their susceptibility to corporate interference. The naïve solution of staffing MSAs with honest administrators is impractical because supervised industries exert pervasive and well-documented pressure on individual supervisors. Further, even stalwart employees have to abide by the orders of politically appointed agency heads with their own agendas and interests. Not to mention that the convoluted enforcement processes designed to hide loopholes weakens regulatory efficacy and hinders sanctions.

3.1 MSAs Are Institutionally Prone to Capture

Periodically, a catastrophic crisis or the advent of a new technology prompts political authorities to set up (a host of) new regulators equipped with vaguely defined resources and powers that amount to a fraction of the riches and clout of the corporate behemoths they are supposed to keep in check. Probably, a dose of parliamentary and public scrutiny could keep the agencies in line. However, the interest invariably fades away within a short time. Especially today, in the age of permanent crisis, attention is swiftly shifted to the next big thing. Yet some people's eyes stay on the ball: those of industry lobbyists. And this is the perfect time for what economists define as the "capture" of market regulators and overseers. As Stigler (1971, p. 3) argued, the result of this process is that, "as a rule", the competent agency "is acquired by the industry and is designed and operated primarily for its benefit" rather than the common good. Still, the AI Act seems to completely ignore this simple fact. One reason is that no entry test can select individuals specifically for qualities like integrity and honesty. But a deeper one appears in the functioning of MSAs, which many are often unaware of. The overseen corporations engage with MSAs through institutionalised channels enshrined in the law of the land and more or less transparently. Nevertheless, industry also interacts with individual agency employees on all levels, including field inspectors responsible for enforcement decisions. This gradual takeover of the supervisory agency by Big Corp makes captured agencies worse than a lack of supervision as they grant the industry a veil of legitimacy and pervert liberal values. The AI Act facilitates capture by giving the Board an outsized role, somewhat like a shepherd putting the wolf in charge of repairing the fence to keep it away the sheep.

In particular, Bulgarian agencies have an abysmal track record when it comes to countering powerful corporate interests. Possibly, the best researched cases in this area involves Big Tobacco. Despite operating in a relatively low-tech industry, tobacco companies have managed to outsmart and capture several key overseeing and regulatory agencies. Basically, these lobbyists' ability to capture regulators and policymakers is felt distinctively in every corner of government, including the Ministry of Economy and Industry, the Ministry of Finance, and the Customs Agency (Gavrailova et al., 2022). For instance, there clearly is room for enforcement given the many sugar-coated and one-sided interviews and articles depicting tobacco products as not-so-harmful or even harmless (Antonov et al., 2020). But the MSA in charge of electronic media (Săvet za elektronite medii, SEM) has turned a blind eye to key outlets' promotion of pro-tobacco talking points and its correlation with sponsorship agreements (Antonov & Barova, 2019). Moreover, the influence of Big Tobacco extends to the MSA that could push to increase excise duties on tobacco products or enforce stricter advertising bans (Antonov & Ivanov, 2020). Even a notoriously noxious industry like Big Tobacco can thus capture public health authorities, the mainstream media, and its regulator in Bulgaria. One can only wonder what the AI oligopoly will accomplish with its technological edge and unlimited resources.

3.2 MSAs Are Politically Prone to Capture

Lawmakers create market overseeing and regulating agencies to rein in powerful forces in society (usually corporations), which yield vast economic and technological benefits, but are also vulnerable to abusing their power. Yet, in the words of a former employee of the US Environmental Protection Agency, the reality is that:

> *employees soon learn that drafting and implementing rules for big corporations means making enemies of powerful and influential people. They learn to be 'team players,' an ethic that permeates the entire agency without ever being transmitted through written or even oral instructions. People who like to get things done, who need to see concrete results for their efforts, don't last long. They don't necessarily get fired, but they don't advance either; their responsibilities are transferred to others, and they often leave the agency in disgust. The people who get ahead are those clever ones with a talent for procrastination, obfuscation, and coming up with superficially plausible reasons for accomplishing nothing.* (Latham, 2012)

The system's fatal flaw is the same on both sides of the Atlantic: the MSA's head is typically a political appointee. And, even if confirmed by the legislature, they are selected based on the preferences of those in power. Hence, MSAs can end up functioning as extensions of the executive branch and operate within political constraints. MSAs' undue politicisation magnifies the power Big Corp holds over the political system. For starters, it is well known that corporate money funds (and bribes) regional governors, heads of government, and members of both the national and EU parliament. Further, the threat of a corporation off- or near-shoring makes governors and prime ministers reluctant to enforce laws. In the arm-wrestling contest between public and corporate interests, any discretionary authority granted to the MSAs' heads benefits the current political majority and effectively serves the vested interests backing it.

In the case of Bulgaria, leading political figures are barely concealing their wish to influence independent authorities through political appointments. A recent instance involved the party Prodălzhavame Promyanata (PP, lit. "We continue the change"), which not long ago joined the liberal Renew in the European Parliament. The leaders of PP, Kiril Petkov and Asen Vasilev, were caught on tape saying "These are our people who are approved by the embassy" when referring to the heads of independent organisations (quoted in Dimitrova, 2023). Similarly, several analysts voiced the opinion that PP's government partner, the centre-right GERB (lit. "flag"), called snap elections because it wanted to appoint its people ("Izbori 2024 – Rezultatăt", 2024, 4:27:00 onwards). And even during the PP-GERB non-coalition's (or sglobka in Bulgarian) tenure, the two parties constantly clashed over the issue of appointments, leading to calls to establish an ad-hoc "council" (Arabadzhiev, 2023).

4 A Policy Blueprint for the MSA Ecosystem

In summary, the AI Act's attempt to bring about the total maximum harmonisation of AI governance has caused the neglect of national specificities as most evident in the framework laid out for MSAs. Although the regulation establishes a veritable ecosystem with seemingly vast powers, far-reaching prerogatives and adequate resources, this vision is likely to remain a mirage. In fact, several provisions are efficiency dampening – rather than enhancing – and seem to disfavour especially poorer peripheral states that could gain enormously from AI, like Bulgaria. As explored in depth above, the AI Act pretends to give freedom to the MSs to choose their own MSA/s, only to nudge them toward disputable choices such as existing financial and data protection regulators. And it does so without regard to these institutions' actual ability to achieve their existing goals, let alone assessing the competency concerning AI. Still, one may justify this attempt in the name of simplifying international cooperation: two data protection regulators may coordinate more easily than two agencies operating in different fields or even brand-new ones. However, one of the least well-designed features of the AI Act's MSA ecosystem is the fragmentation of the MSA-coordination infrastructure. Moreover, these agencies' real powers are limited by complex appeal procedures, even for purely national actions. Vitally, the Commission has the ability to raise objections against enforcement decisions and decide on their merits. Essentially, MSAs will be exposed to the pressures coming from national policymakers and bound by the Commission. This gives the special interest groups behind EU and national political majorities two channels to interfere with market surveillance and enforcement. Finally, adding insult to injury, the AI Act provides a vast remit and a key role in cross-border investigation to a collection of politically appointed lobbyists called the "Board". All in all, this is the perfect recipe for regulatory capture.

Still, this state of affairs is not beyond remedy. On the contrary, a rethinking of the MSA ecosystem could have beneficial effects on neighbouring policy areas in which the EU has also struggled to affirm itself as a leading global power. In particular, it is only by looking cross-sectionally that the MSA ecosystem can live up to its original vocation as the defender of liberal values against the AI oligopoly. All efforts should therefore focus on the following three policy areas of varying scope and increasing geopolitical relevance.

4.1 Empowering and Incentivising Capable and Honest Staffers

First, the EU should take stake of the political economy of market surveillance and regulation and focus on staff's honesty and skills as a counterweight to lobbyists' attempts to undermine enforcement. Here, there are two dimensions that ought to be at the centre of rigorous hiring practices.

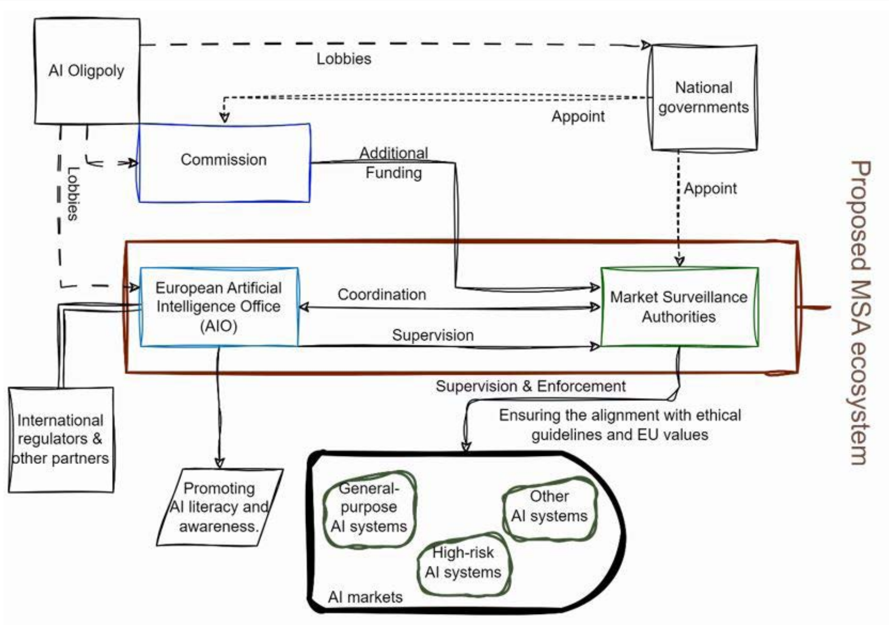
On one hand, recruiters must ensure that MSAs' staffers are not only qualified, but able to sustain the pressure of constant upskilling. In this sense, the bulk of the effort must be supporting poorer MSs by providing up-start and ongoing training as well as professional development opportunities to keep agency staff updated on the latest advancements in AI technologies, ethics, and to exchange best practices with their colleagues from abroad. A very important step in this direction would be to follow the path of the UK, which has "invested over £2.5 billion in AI since 2014" (Donelan, 2023) and allocated "£100 million to help realise new AI innovations and support regulators' technical capabilities", including "£10 million to jumpstart regulator's AI capabilities" (Donelan, 2024). Equivalently, lawmakers should create an EU fund for upskilling national regulators allocated based on actual needs to pay for tech and language training as well as staff mobility and internationalisation.

On the other hand, hiring should also be conditional on candidates possessing a strong ethical foundation. A judgment here should be based on an assessment of their commitment to public service and liberal values as well as investigations of potential conflicts of interests. Efforts should also be made to recruit individuals with a vested interest in effective enforcement. To reap all the benefits of having a stalwart and ethical staff, there should however be an EU rule to incentivise whistle-blowers with a provision like those found in the USA's Federal Civil False Claims Act. Among other benefits, whistle-blowers can expect monetary rewards that, while sounding appealing, amount to a fraction of the savings from a reduction in corruption and the gain from improved enforcement.

4.2 Streamlining the MSA Ecosystem

At the moment, the MSA ecosystem is cluttered with a number of actors that are entitled to extraordinarily pervasive prerogatives. This does not simply contribute to making the transfer of knowledge and best practices between MSAs more difficult, but it also gives the AI oligopoly and political majorities, on both the national and EU levels, multiple ways to pressure MSAs. These two issues can be solved by streamlining the MSA ecosystem and putting an enhanced Artificial Intelligence Office (AIO) in charge of a new coordination framework.

Figure 2: A streamlined MSA ecosystem for effective enforcement


AI Oligpoly
Lobbies
National governments
Appoint
Commission
Additional Funding
Lobbies
Appoint
Proposed MSA ecosystem
European Artificial Intelligence Office (AIO)
Coordination
Supervision
Market Surveillance Authorities
Supervision & Enforcement
Ensuring the alignment with ethical guidelines and EU values
International regulators & other partners
Promoting AI literacy and awareness.
General-purpose AI systems
Other AI systems
High-risk AI systems
AI markets

Source: Author original elaboration

Currently, the AI Act assigns minor functions to the AIO, under the Commission. Primarily, it is another body tasked with coordinating the regulation's implementation across the EU, together with the Commission and the Board. In addition, the AIO supervises and monitors the compliance of general-purpose AI systems with the Act's requirements, overseeing the development, deployment and use of such systems (AI Act, 2024, Nº 161–162). Moreover, the AI Office supports member states and stakeholders in the AI ecosystem by offering technical assistance, sharing best practices, and facilitating knowledge exchange to promote the development of safe AI. It also operates regulatory sandboxes for the real-world testing of general-purpose AI technologies to understand their implications and develop appropriate responses with limited enforcement authority to investigate non-compliance and impose fines. Finally, the Office promotes AI literacy and public awareness about AI technologies' benefits and risks, educating stakeholders about their rights and obligations under the AI Act. Still, the AIO is not the MSA for EU bodies and agencies because it is hierarchically subordinated to the Commission, with the Data Protection Supervisor having been preferred.

> ”
> The artificial intelligence office should take on the functions of coordination between national market surveillance authorities.

Yet, the subordination to the Commission is not a good reason for overburdening an organisation already busy dealing with enforcing the GDPR with a radically different role. Further, national MSAs are also liable to see all their decisions overridden by the Commission. The impression that the Commission is the apex predator in the MSA ecosystem is thus unmissable. It would instead be preferable to redistribute competences amongst these organisations differently. The first step would be to make the AIO independent of the Commission, which either way acts more as a parent than a partner with most Union agencies (Vestlund, 2015). This would not be the first time that a brain child of the Commission acquired its own physiognomy, the same happened to the European Medicines Agency (EMA). Originally conceived as the "European Agency for the Evaluation of Medicinal Products" (Directive 93/41/EEC, 1993) and established in 1995, the agency has become increasingly independent over the decades since its transformation to the EMA (Regulation 726/2004/EC, 2004). At the moment, the EMA is directly accountable to European citizens, 'autonomous' in that its decision-making process is shielded from political or commercial influence, and has its own budget (Regulation 2019/5/EU, 2018, arts. 3, 5, 9, 16). Thus, the same transformation could involve the AIO, allowing it to take on the role of MSA for EU institutions. Symbolically this would also show the MSs that creating a new body for the function of MSA is not only allowed, but practicable. In addition, the newly independent, better funded and properly staffed AIO should take on the functions of coordination between national MSAs, including for cross-border investigations. This would rationalise the messy distribution of competences in this area between the Commission, the Board and the AIO. Moreover, the AIO should become the body tasked with deciding on the objections to national MSAs' enforcement at first instance, removing the judge-and-jury power the Commission presently enjoys. To ensure that the rule of law is respected, these decisions should naturally be appealable at the Court of Justice.

To make sure that this more powerful AIO is both the EU MSA and the only locus of coordination between MSAs, the Board should be abolished. Besides consolidating the AIO as the guarantor of MSAs' refractivity to a bland attempt of infiltration or capture, removing the Board would deprive the AI oligopoly of a path by which to channel their lobbying efforts. The reservoir of expertise that belonged in the Board could still be made accessible to the AIO, and via it to all the MSAs, as a mere consultative body with no real saying power. Meanwhile, the AIO should take a page out of the EMA's book and implement public hearings as a tool to fight against special interests and support democratic values and practices (Wood, 2022).

4.3 Granting MSAs Real Powers: A Path Forward

Once the national MSAs are empowered and the coordination between them is efficiently centralised in a newly-independent AIO, it would be time to endow the MSAs with real powers. This empowerment is crucial for ensuring that the AI Act's regulatory framework is not simply robust in theory but also effective in practice. The following recommendations outline the steps needed to achieve this goal.

To address the tension between a maximally harmonised regulation and national enforcement, MSAs require a degree of policy space to allow for immediate, effective action. Specifically, MSAs should be granted the authority to immediately take valid nationwide enforcement decisions. Such enforcement powers should be particularly sweeping when it comes to defending fundamental rights and civil liberties, like the right to privacy, free speech, and the complementary right to be heard. These rights are often the first to be compromised in the face of unchecked technological advancements which rapidly become a "weapon of repression" (Ünver, 2024). This means it is vital that MSAs have the capability to intervene decisively in cases where these rights are at stake. Enforcement decisions should only be subject to review by the AIO if there are timely and reasoned objections. And the only actors allowed to call the AIO into question are the Commission and other MSAs that consider themselves (or businesses under their jurisdiction) affected negatively by the decision. This would make sure that MSAs can act swiftly to address issues without being bogged down by bureaucratic delays while still guaranteeing the providers and deployers access to legal remedies if they do not wish to cooperate with the MSA.

> ❞ Market surveillance authorities should be placed at the center of an intergovernmental framework for public investment and acquisitions in AI.

In addition, MSAs should have the ability to name and shame companies that do not cooperate or repeatedly infringe upon civil liberties and fundamental rights. This public disclosure mechanism can serve as a powerful deterrent against non-compliance in a high-risk AI system and may encourage greater accountability among providers and deployers. Importantly, this naming and shaming should also apply to public sector entities, including law enforcement and civil protection agencies that violate the terms of emergency authorisation. It is critical that these institutions should not be allowed to authorise the deployment of non-conforming high-risk AI systems without the presence of a properly qualified representative of the national MSA. This provision ensures that even in situations deemed urgent, civil liberties and liberal values are protected.

By implementing these measures, MSAs can act as a veritable (and well-armed) David in the fight against the encroaching Goliath that is the AI oligopoly. Empowered MSAs can provide a robust check on the power of major AI companies, assuring that the deployment of AI technologies remains within the bounds of the law and respects fundamental rights.

To further strengthen their role, MSAs should be placed at the centre of an intergovernmental framework for public investment and acquisitions in AI. This framework would support the development and deployment of AI systems that are bred within the EU, particularly those produced on the periphery. This could occur by giving priority to the applications for the conformity, sandboxing and real-world testing of autochthonous systems on the national MSA levels, while a special committee with the AIO would oversee the awarding of joint research activity and grants paying special regard to spatial and social inequalities. By prioritising investments in EU-bred systems, this framework would help to build a more resilient and independent AI ecosystem within the Union and foster cohesion in favour of peripheral MSs with significant comparative advantages in AI, like Bulgaria.

The AIO should also act as an advisory body to an intergovernmental council composed of national MSA staffers and tasked with earmarking national funds for public-sponsored mergers and acquisitions in the AI market. Functioning on the basis of unanimity, this council would naturally serve the interests of peripheral countries while also strategically directing resources to AI projects that align with the EU's values and policies, including the promotion of AI systems enhancing public goods like healthcare, education and upskilling. By linking public investment with these priorities, the EU can ensure that the benefits of AI are broadly shared and contribute to the well-being of all its citizens.

5 Conclusion

In conclusion, the AI Act's framework for national supervisory authorities (MSAs) exposes considerable vulnerabilities, especially for smaller and less affluent EU member states. While the legislation aims to create a harmonised approach to AI regulation, it inadvertently nudges countries towards choosing existing institutions, often ill-equipped to handle the complex and ever-evolving demands of AI oversight. This rigidity, combined with the substantial financial and technical requirements imposed by the Act, places an undue burden on poorer countries, potentially compromising the effectiveness of market surveillance.

The Act's dispersed coordination responsibilities further undermine the MSAs' efficiency. The

involvement of multiple bodies without a clear, centralised forum for cooperation creates a fragmented regulatory landscape, reducing the overall effectiveness of AI oversight. In addition, the potential for regulatory capture by the AI oligopoly is exacerbated by the influence of lobbyists and the Commission's overarching authority, which can override national MSAs' decisions. The focus on Bulgaria, a peripheral country with largely untapped AI potential, underscores these challenges, illustrating how weak institutional frameworks can lead to the capture of regulatory agencies by powerful corporate interests. This issue is compounded by political interference, with agency heads often being political appointees, which entrenches special interests.

To address these challenges, the policy paper proposes several key recommendations. First, it calls for the establishment of an EU fund to upskill national regulators, assuring they have the technical and ethical foundations required to withstand industry pressures. Incentives for whistle-blowers and enhanced judicial oversight are also recommended to bolster the integrity and effectiveness of MSAs. Second, the paper calls for a streamlined MSA ecosystem, centred around an upgraded and independent European Artificial Intelligence Office (AIO). This body would coordinate national MSAs, oversee cross-border investigations, and act as the primary enforcer of AI regulations for EU institutions, replacing the politically vulnerable Board. Finally, the paper suggests granting MSAs real enforcement powers, allowing them to take immediate, binding actions to protect civil liberties and fundamental rights. This empowerment should be complemented with a public disclosure mechanism to name and shame non-compliant entities and an intergovernmental framework for public investment in AI, prioritising EU-bred systems and nourishing cohesion among member states.

These measures aim to create a robust and effective regulatory environment capable of counterbalancing the power of the AI oligopoly and safeguarding the EU's liberal values. By addressing the specific needs and vulnerabilities of peripheral countries like Bulgaria, the EU can ensure a more equitable and resilient approach to AI governance, promoting innovation while protecting the rights and interests of its citizens.

REFERENCES

- Antonov, P., & Barova, V. (2019). Липсващото С: Преглед на спонсорството от страна на тютюневата индустрия в България [The missing C: An overview of the industry patronage industry in Bulgaria] (Analytical report, pp. 1–8) [Policy recommendations]. BlueLink Foundation. https://www.bluelink.net/files/attachments/missing-s_report_2019.pdf
- Antonov, P., Barova, V., & Vakhrusheva, K. (2020). Фокусът с нагреваемите тютюневи продукти в медиите: Изследване на съдържанието и честотата на изгодните за индустрията послания в медиите и причините за тях [Focusing on heated tobacco products in the media: A study of the content and frequency of industry-friendly messages in the media and the reasons behind them] (Analytical report, pp. 1–20) [Policy recommendations]. BlueLink Foundation. https://www.bluelink.net/files/attachments/focus-on-htp-in-media.pdf.pdf
- Antonov, P., & Ivanov, H. (2020). Корона и тютюн [Coronavirus and Tobacco] (Report, pp. 1–17) [Policy brief]. BlueLink Foundation. https://www.bluelink.net/files/attachments/covid-19_smoking_and_mortality_report_0.pdf
- Arabadzhiev, I. (2023, July 23). ПП-ДБ: Формираме Борд за назначенията с ГЕРБ-СДС [PP-DB: We are forming an Appointments Board with GERB-SDS] (D. Radeva, Interviewer) [Euronews Bulgaria]. https://euronewsbulgaria.com/news/17076/iskren-arabadzhiev-pp-db-formirame-bord-za-naznacheniyata-s-gerb-sds
- Council Directive 93/41/EEC of 14 June 1993 Repealing Directive 87/22/EEC on the Approximation of National Measures Relating to the Placing on the Market of High-Technology Medicinal Products, Particularly Those Derived from Biotechnology, Pub. L. No. 93/41/EEC, 1 (1993). https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX%3A31993L0041%3AEN%3AHTML
- Dimitrova, R. (2023, May 26). Цитатите от записа на Василев: От одобрени от посолство хора в службите до изпиране на Борисов [Quotes from Vassilev's recording: from embassy-approved stagger in the security services to Borissov's money laundering] [News agency]. Dnes. https://dnes.dir.bg/na-fokus/tsitatite-ot-zapisa-na-vasilev-ot-odobreni-ot-posolstvo-hora-v-sluzhbite-do-izpirane-na-borisov
- Donelan, M. (2023). A pro-innovation approach to AI regulation (Policy Paper 815; White Paper). UK Secretary of State for Science, Innovation and Technology. https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
- Donelan, M. (2024). A pro-innovation approach to AI regulation: Government response (Policy Brief CP 1019; Consultation Outcome). UK Secretary of State for Science, Innovation and Technology. https://www.gov.uk/government/consultations/ai-regulation-a-pro-innovation-approach-policy-proposals/outcome/a-pro-innovation-approach-to-ai-regulation-government-response
- Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Pub. L. No. EO 14110 (2023). https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/

- Gavrailova, M., Geshanova, G., & Antonov, P. (2022). Токсична зависимост: Как индустрията формира в своя полза политиките за ограничаване употребата на тютюневи изделия [Toxic Addiction: How the Industry Shapes Policies to Curb Tobacco Use in Its Favour]. https://coalicia.bezdim.org/wp-content/uploads/2022/03/2022-toksichna-zavisimost-doklad.pdf
- Georgieva, I., Timan, T., & Hoekstra, M. (2022). Regulatory Divergences in the Draft Ai Act: Differences in Public and Private Sector Obligations. Publications Office. https://doi.org/10.2861/69586
- Hankins, E., Nettel, P. F., Martinescu, L., Grau, G., & Rahim, S. (2023). 2023 Government AI Readiness Index (Policy Brief 5; pp. 1–53). Oxford Insights. https://oxfordinsights.com/ai-readiness/ai-readiness-index/Oxford Insights
- Hern, A. (2024, February 6). TechScape: Why is the UK so slow to regulate AI? The Guardian. https://www.theguardian.com/technology/2024/feb/06/techscape-uk-artificial-intelligence-government-regulation
- Latham, J. (2012, May 1). Designed to Fail: Why Regulatory Agencies Don't Work. Independent Science News. https://www.independentsciencenews.org/health/designed-to-fail-why-regulatory-agencies-dont-work/
- Marcelli, S., Haefele, M., Hoffmann-Burchardi, U., Dennean, K., Monnet, A., Stiehler, A., & Gantori, S. (2024). Sizing and seizing the AI investment opportunity (Chief Investment Office's Daily Updates, pp. 1–6) [Briefing]. UBS Wealth Management. https://www.ubs.com/global/en/wealth-management/insights/chief-investment-office/house-view/daily/2024/latest-11062024.html
- Nobel Prize Outreach. (1982). The Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel 1982. Nobel Prize. https://www.nobelprize.org/prizes/economic-sciences/1982/summary/
- Plotinsky, D., & Cinelli, G. M. (2024). Existing and Proposed Federal AI Regulation in the United States (Insight) [Policy brief]. Morgan, Lewis & Bockius. https://www.morganlewis.com/pubs/2024/04/existing-and-proposed-federal-ai-regulation-in-the-united-states
- Regulation (EC) No 726/2004 of the European Parliament and of the Council of 31 March 2004 Laying down Community Procedures for the Authorisation and Supervision of Medicinal Products for Human and Veterinary Use and Establishing a European Medicines Agency, Pub. L. No. Regulation (EC) 726/2004 (2004). https://eur-lex.europa.eu/eli/reg/2004/726/oj
- Regulation (EU) 2019/5 of the European Parliament and of the Council of 11 December 2018 Amending Regulation (EC) No 726/2004 Laying down Community Procedures for the Authorisation and Supervision of Medicinal Products for Human and Veterinary Use and Establishing a European Medicines Agency, Regulation (EC) No 1901/2006 on Medicinal Products for Paediatric Use and Directive 2001/83/EC on the Community Code Relating to Medicinal Products for Human Use (Text with EEA Relevance), Pub. L. No. Regulation (EU) 2019/5, 004 OJ L (2018). http://data.europa.eu/eli/reg/2019/5/oj/eng
- Regulation of the European Parliament and of the Council Laying down Harmonised Rules on Artificial Intelligence, Pub. L. No. P9_TA(2024)0138, 1 (2024). https://www.europarl.europa.eu/doceo/document/TA-9-2024-0138_EN.pdf
- Sitaraman, G., & Narechania, T. (2023). Antimonopoly Tools for Regulating Artificial Intelligence (Vanderbilt Policy Accelerator, pp. 3–54) [Policy brief]. Vanderbilt University. https://cdn.vanderbilt.edu/vu-URL/wp-content/uploads/sites/412/2023/10/06212048/Narechania-Sitaraman-Antimonopoly-AI-2023.10.6.pdf.pdf

- Stigler, G. J. (1971). The Theory of Economic Regulation. The Bell Journal of Economics and Management Science, 2(1), 3–21. https://doi.org/10.2307/3003160
- Telarico, F. A. (2024). Back to the future: Can Bulgaria become Europe's AI hub? (E. N. Tomaževič, D. Ravšelj, & A. Aristovnik, Eds.; pp. 156–175). European Liberal Forum.
- Ünver, H. A. (2024). Artificial intelligence (AI) and human rights: Using AI as a weapon of repression and its impact on human rights (Policy Paper PE 754.45; In-Depth Analysis, pp. 1–119). European Parliament - Subcommittee on Human Rights (DROI). https://www.europarl.europa.eu/RegData/etudes/IDAN/2024/754450/EXPO_IDA(2024)754450_EN.pdf
- Veale, M., & Borgesius, F. Z. (2021). Demystifying the Draft EU Artificial Intelligence Act – Analysing the good, the bad, and the unclear elements of the proposed approach. Computer Law Review International, 22(4), 97–112. https://doi.org/10.9785/cri-2021-220402
- Vestlund, N. M. (2015). Exploring the EU Commission-Agency Relationship: Partnership or Parenthood? In M. W. Bauer & J. Trondal (Eds.), The Palgrave Handbook of the European Administrative System (pp. 349–365). Palgrave Macmillan UK. https://doi.org/10.1057/9781137339898_20
- Wood, M. (2022). Can independent regulatory agencies mend Europe's democracy? The case of the European Medicines Agency's public hearing on Valproate. The British Journal of Politics and International Relations, 24(4), 607–630. https://doi.org/10.1177/13691481211054319
- Избори 2024 – Резултатът [Election 2024: The result]. (2024, June 9). [16:9]. In Po sveta i u nas. BNT 1. https://bntnews.bg/news/izbori-2024-rezultatat-1281386news.html

Chapter 7

The Regulatory Dimension of Artificial Intelligence in Romania

Melania-Gabriela Ciot

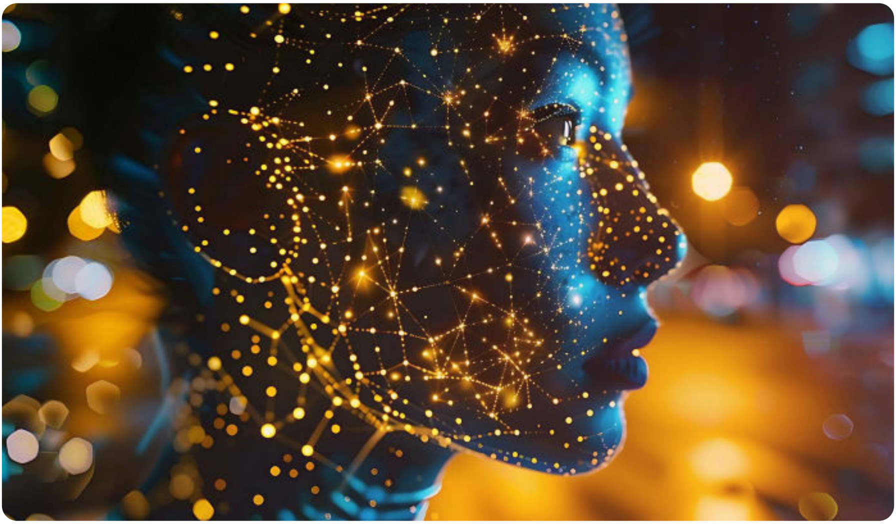
1 Introduction

The European Union succeeded in 2023 to regulate the use of artificial intelligence (AI) by initiating the AI Act, the first AI law globally, which is seen as the normal step forward, after the elaboration of the digital strategy. In this way, the EU's capacity in new digital technologies will be strengthened, better policies will be envisioned, that will support the green and digital transition and will consolidate the instruments for reaching the targets of climate neutrality for 2050. The European institutions are fully committed to creating the context and tools for the digital transformation, as the Digital Decade Programme announced, focusing on the integration of the digital single market while improving the use of AI, digital skills and innovation.

However, the success of AI development in the EU will depend on the member states' capacity and ability to implement their national strategy. This explains why the legislative harmonisation of the AI Act on the national level will have to identify the commonalities and differences in European legislation and create the possibilities for adjustments and improvements. For Romania, as an example, important steps have already been taken in the AI sector either when speaking about its introduction in different sectors of activities or the creation of a necessary regulation framework. It is also worth mentioning that the regulatory dimension for AI in Romania is in the process of elaboration and significant advancements have already been made.

2 Newest Approaches to AI on the European Level

The EU's AI approach is considered one of the most innovative on the global scale given that it is based on research and innovation, respect of fundamental human rights, and regulation of the use of AI (European Commission, 2024a). The preoccupation and initiatives of EU in this regard began in March 2018 with the initiation of European AI Strategy and establishment of the first AI experts' group, appointed by the European Commission, tasked with elaborating the ethics guidelines for AI (European Commission, 2018a), and continuing till the present with the latest amendments to the EuroHPC Regulation (European Commission, 2024b).



To fully understand how AI has evolved, it is essential to delve into the chronological progress of its development and emphasise the milestones and breakthroughs that have collectively shaped the current state of AI – from the early conceptual theories and rudimentary machines to the sophisticated algorithms and applications:

Table 1: Important steps in adopting the AI Act

Year	Month	
2018	March	Initiation of the expert group for the elaboration of AI ethics guidelines
	April	Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe, COM/2018/237 final
		Commission staff working document. Liability for emerging digital technologies, COM/2018/237 final
		Declaration of cooperation on artificial intelligence – Romania joined the initiative from May 2018
	June	Launch of the EU's AI Alliance
	December	Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Coordinated Plan on Artificial Intelligence, COM/2018/795 final
2019	April	Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Building Trust in Human Centric Artificial Intelligence [COM(2019)168]
	June	The first European AI Alliance Assembly
		Policy and Investment Recommendations on AI, elaborated by the European AI Alliance and addressed to the European Commission and member states
2020	February	White Paper on Artificial Intelligence: A European approach to excellence and trust, COM(2020) 65 final
	July	Final assessment of the High-Level Expert Group on Artificial Intelligence (AI HLEG) for Trustworthy Artificial Intelligence (ALTAI)
	October	Second European AI Alliance Assembly

2021	April	Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European approach to Artificial Intelligence, COM/2021/205 final
		Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, COM/2021/206 final
	June	Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council, COM/2021/346 final
	November	Council of the EU: Presidency compromise text on the AI Act
		Opinion of the European Economic and Social Committee on Regulation of the AI Act
	December	Committee of Regions' Opinion on the AI Act – European Approach to Artificial Intelligence – Artificial Intelligence Act (revised opinion)
		Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence, (CON/2021/40) (2022/C 115/05)
2022	June	Spain's pilot for a Regulatory Sandbox on Artificial Intelligence (AI)
	September	Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive), COM/2022/496 final
	December	Council of EU: Adoption of a general approach to the AI Act
2023	June	European Parliament Negotiations on the AI Act
	December	Political agreement of the European Parliament and the Council on the Artificial Intelligence Act
2024	January	An AI innovation package to support artificial intelligence startups and SMEs
	February	European AI Office
		Approval of the world's first comprehensive rulebook for artificial intelligence

Source: European Commission, 2024

An analysis of the milestones presented in Table 1 shows the intense activity of European institutions, especially the European Commission, regarding the regulation and establishment of rules and guidelines for AI's trustworthy use and also the preoccupations with the ethical dimensions of AI, that could be considered to be a landmark of the European Union, known as a promotor and defender of human rights and of its European values. In just a period of 6 years, through comprehensive negotiations the EU delivered the AI Act and launched the AI Pact and initiated the first regulatory proposals on the approach to AI (by the European Commission, between 2018 to 2021). Other European institutions, such as the Council of EU, Committee of Regions, European Central Bank, European Economic and Social Committee and European Parliament, expressed their opinions on the AI Package, improving its initial form (2022 and 2023), consolidating a solid base for future developments. Another interesting approach to the initiatives the EU has taken for regulating the use of AI belongs to a prestigious European think-thank group – the European Council of Foreign Relations, which introduced the term "digital decolonisation" (Klein, 2020) while referring to an initiative of the European Commission (White Paper on Artificial Intelligence, European Data Strategy and Shaping Europe's Digital Future), seen as an inclusive process through which the European Union is trying to escape from other global actors for technologies, such as the USA and China. In practice, these regulations of AI may be interpreted as the "EU's aspiration for technological sovereignty" (Ortega Klein, 2020).

Due to changes in the international context, the EU saw some delays in the implementation of its initiatives, mainly concerning the elaboration of a suitable framework for innovation and investments to consolidate the economic competitive advantage, strategic autonomy and resilience (Renda, 2024). Trustworthy use of AI is the defining element of the European approach, which seeks to emphasise the importance of research and innovation, as well as European values and principles when it comes to characterising the endeavours made and projected for AI use. Renda (2024) considers that in terms of competitiveness, strategic autonomy and technological sovereignty, the EU made AI one of its main priorities and, because of its impact on European economies and societies, democracies and services, AI should be treated from the perspectives of its vulnerability to external pressure that could dimmish its achievements in terms of the economy, standards, principles and regulations.

On a global level, the progress in AI fields is immense and was accelerated in 2023 (Maslej et al., 2024). Advancement with the development of GPT-4, Gemini, Claude 3, or the AI products developed by companies, for the public use reveals the following future tendencies: (1) continuing improvements of technology; and (2) constraints on the use of AI due to the limitations of technology (Maslej et al., 2024), but also from the regulation. Considering the policymaking process, AI regulation has never been higher than in the past year (2023). Further, there are many concerns regarding the use of AI for electoral purposes or to generate deepfakes. The Standford AI Index (Maslej et al., 2024) highlights several crucial characteristics of AI development, taking into consideration the current state of AI, key trends, and advancements that define the field. It examines the rapid growth in AI research, the increasing sophistication of machine learning models and expanding applications of AI in various industries:

Table 2: Characteristics of AI development in 2023

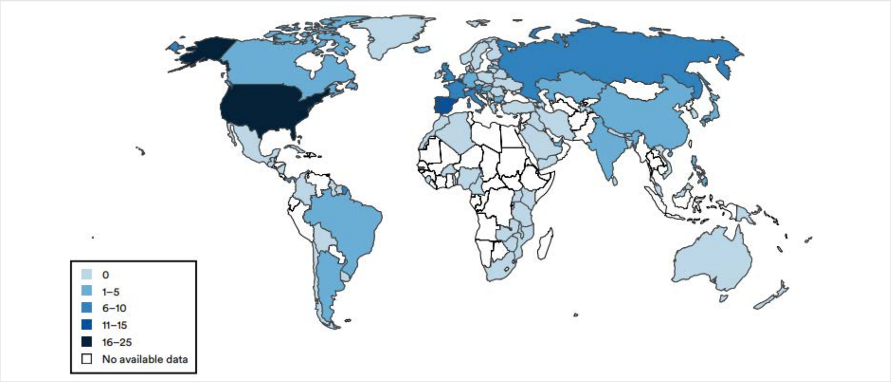
Characteristic	Description
AI outperforms humans in certain tasks, but not all	AI has exceeded human performance in various benchmarks, for example image classification, visual reasoning, or English comprehension). However, it still lags in more complex tasks like competition-level mathematics or strategic planning.
The industry continues to lead in cutting-edge AI research	In 2023, the industry developed 51 notable machine learning models, while academia produced only 15
Frontier models are becoming significantly more expensive	The training costs for state-of-the-art AI models have soared to unprecedented levels such as, for example, OpenAI's GPT-4 required an estimated USD 78 million in computing costs, while Google's Gemini Ultra demanded USD 191 million for computing.
The USA surpasses China, the EU, and the UK as the primary source of top AI models	In 2023, U.S.-based institutions produced 61 notable AI models, easily surpassing the European Union's 21 and China's 15
There is a significant lack of robust and standardised evaluations for assessing the responsibility of large language models (LLMs)	Major developers like OpenAI, Google and Anthropic predominantly evaluate their models against various responsible AI benchmarks; this complicates efforts to consistently compare the risks and constraints of top AI models
Investment in generative AI has surged dramatically	Investment in generative AI has experienced a considerable surge to reach USD 25.2 billion. Key players in the generative AI sector, such as OpenAI, Anthropic, Hugging Face, and Inflection, reported substantial fundraising rounds

AI enhances worker productivity and leads to higher quality output	Several studies have evaluated AI's effect on labour, indicating that AI enables workers to complete tasks more efficiently and increase the quality of their output
The progress in science has increased exponentially	The introduction of increasingly impactful science-related AI applications – from AlphaDev, which enhances algorithmic sorting efficiency, to GNoME, which streamlines the process of materials discovery
AI regulations in the USA have seen a sharp increase	In 2023, there were 25 AI-related regulations, a notable increase from just 1 in 2016. The total number of AI-related regulations grew by 56.3% in the past year alone.
People worldwide are increasingly aware of AI's potential impact – and more apprehensive about it	According to a survey by Ipsos, the share of people who believe AI will significantly impact their lives in the next 3 to 5 years has risen from 60% to 66% in the past year

Source: Stanford AI Index, 2024

The capabilities of AI have focused the attention of policymakers and decision-makers. Significant AI-related policies have been formulated in the EU and the USA, reflecting the responsibilities of policymakers for elaborating an AI regulatory framework, and for its potential capitalisation (Boyer et al., 2024). The index also underscores the importance of ethical considerations and regulatory frameworks in guiding AI's responsible management, as presented below:

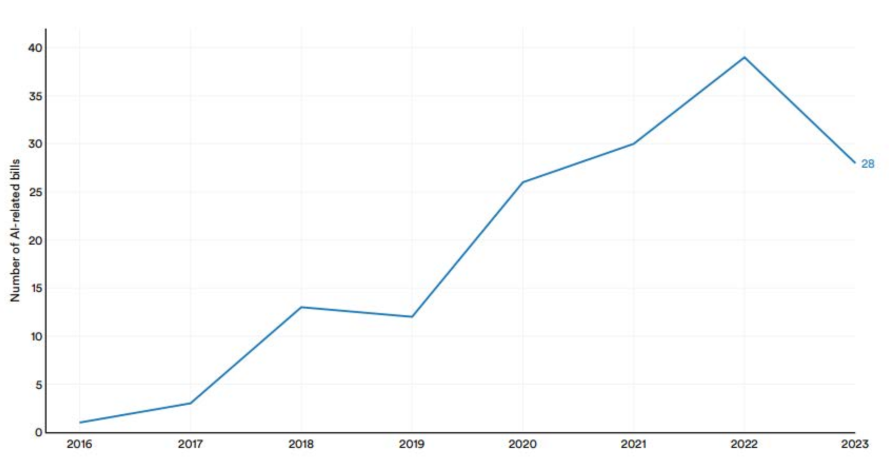
Figure 1: AI-related regulations around the world passed between 2016 and 2023



Source: Stanford AI Index, 2024

The above figure reveals that the majority of AI-related of bills were passed by the USA, the EU (with the notable contribution of Spain) and China. According to the Stanford AI Index (Maslej et al., 2024), 148 AI-related legislative acts were enacted by 128 analysed countries, and 32 countries of these have adopted at least one related AI legislative act, highlighting global recognition of the importance of regulating and guiding the development and deployment of AI. The widespread legislative activity underscores the growing need for a governance framework to address the ethical, social and economic implications of AI technologies, as illustrated below:

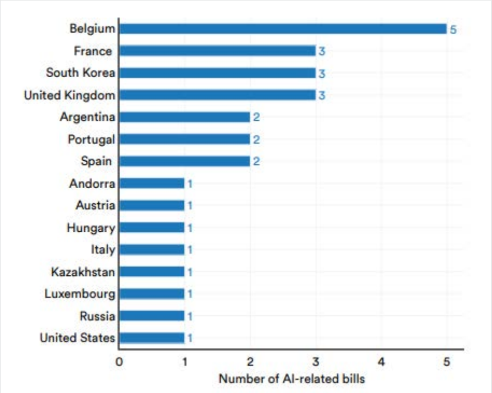
Figure 2: The evolution of AI-related regulations around the world between 2016 and 2023



Source: Stanford AI Index, 2024

As Figure 2 shows, the pinnacle of AI legislative evolution was in 2022, a year that witnessed the global release of nearly 40 legislative acts, marking a significant milestone reflecting the heightened awareness and urgency among nations to address the complexity of AI regulation. In addition, there was a clear surge in AI legislation between 2019 and 2020, with over 20 acts being adopted around the world in this period, that reveals the accelerated pace at which countries are recognising the need for comprehensive legal frameworks to manage the rapid advancements of AI technologies. Observing the geographical areas of adoption of AI-related laws reveals the engagement with crafting AI policies, as Figure 3 below indicates:

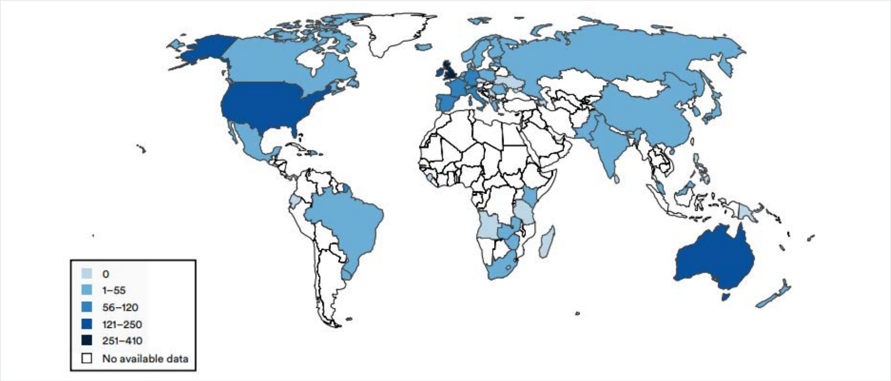
Figure 3: The geographical distribution of AI-related bills transformed into laws



Source: Stanford AI Index, 2024

The figure shows that the EU is leading, being represented by Belgium with 5 laws, followed by France, and then by South Korea and United Kingdom. It could be stated that some regions are more pro-active in governing AI technologies, such as the EU, indicating potential for growth and development in AI governance. Observing these geographical areas for the evolution of AI could lead to a better understanding of the landscape of AI regulation, finding suitable pathways for transforming draft AI laws into legislation, as the following figure shows:

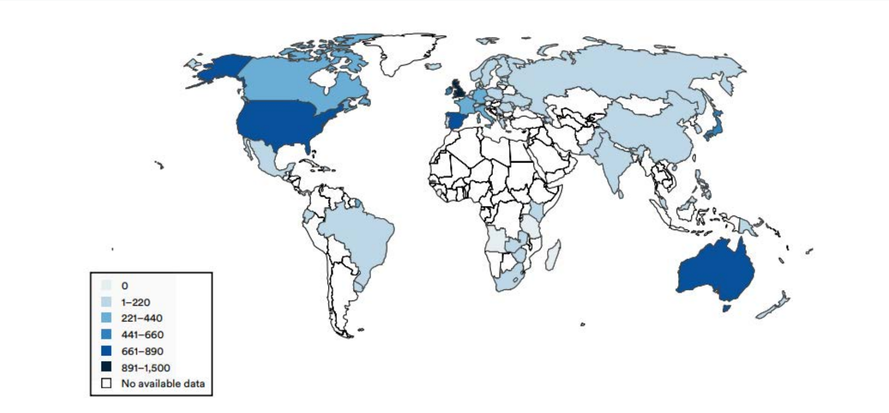
Figure 4: The geographical distribution of AI-related bills transformed into laws



Source: Stanford AI Index, 2024

Figure 4 indicates that the UK dominates, with 405 legislative proceedings followed by the USA (240) and Australia (227) (Maslej et al., 2024), inviting researchers and experts to analyse the contexts and factors that led to this distribution. Several elements could contribute to this distribution, such as technological advancement, infrastructures, investments in AI research and proactive approach to AI governance. This global perspective reveals the importance of the continuous analysis and adaptation of AI policies to the challenges and opportunities generated by global and regional interdependences, as the next figure shows:

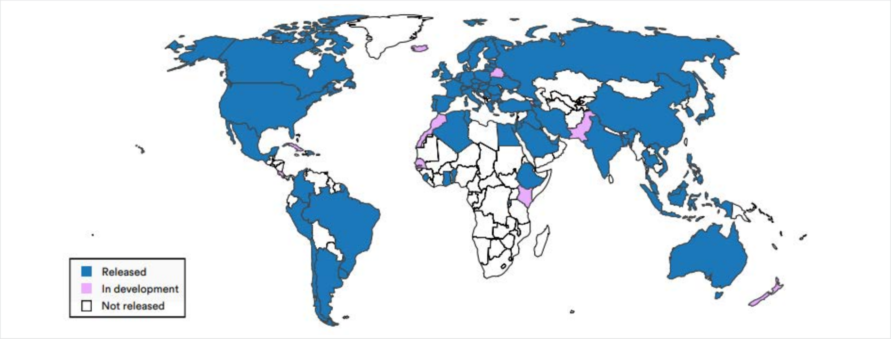
Figure 5: The geographical distribution of AI-related mentions in legislative proceedings



Source: Stanford AI Index, 2024

Figure 5 indicates the dominance of the UK (1,490 mentions), followed by Spain (886 mentions) and the USA (868). Among the European Union, Spain has a very good evolution as regards the establishing of a legislative framework of AI, having in June 2022 initiated the Regulatory Sandbox on Artificial Intelligence (AI), as mentioned previously (European Commission, 2022a). The initiative tries to design a controlled environment for AI innovation testing, while developing the safe use of AI technologies. Similar initiatives are implemented on a global scale as countries around the globe are recognising the need to create suitable frameworks and environments for the testing of AI technologies before they are implemented in societies, as Figure 6 presents:

Figure 6: The geographical distribution of AI-related mentions in legislative proceedings



Source: Stanford AI Index, 2024

When referring to strategies that were adopted for AI use, the Stanford AI Index (Maslej et al., 2024) indicates that Canada was the first country to have elaborated an AI Strategy (March 2017). Today, 75 countries have designed their AI strategies, concentrating in 2019 (when 24 strategies were released), followed by states from the Middle East, Africa or the Caribbean in 2023, as the table below shows.

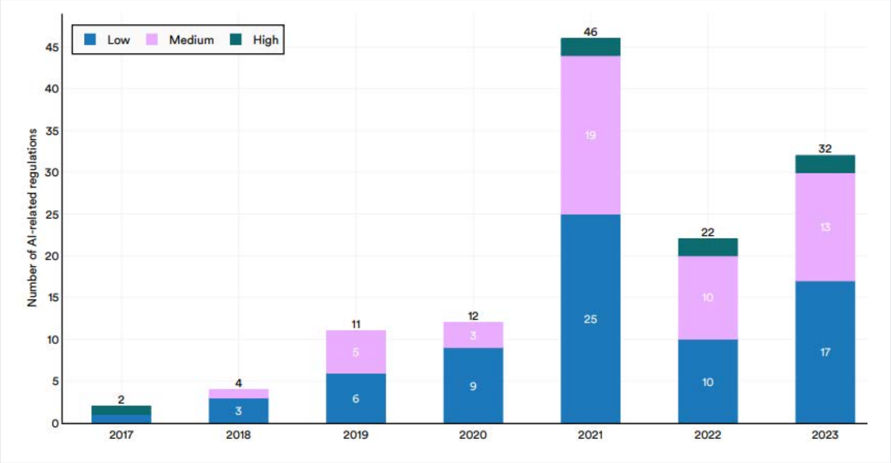
Table 3: Important steps in adopting the AI Act

Year	State
2017	Canada, China, Finland
2018	France, Germany, India, Mauritius, Mexico, Sweden
2019	Argentina, Bangladesh, Chile, Colombia, Cyprus, Czech Republic, Denmark, Egypt, Estonia, Japan, Lithuania, Luxembourg, Malta, Netherlands, Portugal, Qatar, Romania, Russia, Sierra Leone, Singapore, Slovak Republic, United Arab Emirates, United States of America, Uruguay
2020	Algeria, Bulgaria, Croatia, Greece, Hungary, Indonesia, Latvia, South Korea, Norway, Poland, Saudi Arabia, Serbia, Spain, Switzerland
2021	Australia, Austria, Brazil, Hong Kong, Ireland, Malaysia, Peru, Philippines, Slovenia, Tunisia, Turkey, Ukraine, United Kingdom, Vietnam
2022	Belgium, Ghana, Iran, Italy, Jordan, Thailand
2023	Azerbaijan, Bahrain, Benin, Dominican Republic, Ethiopia, Iraq, Israel, Rwanda

Source: Stanford AI Index, 2024

For the EU it is relevant to note the increased number of AI-related legislative acts released in the last few years, especially in 2021 (when 46 acts were approved), and the increase with 11 acts from 2022 and 2023 (Maslej et al., 2024), a fact that demonstrates concentrated activity in this field. The ethical and social aspects of the legislation were the main preoccupations of the European legislator – the European Parliament and now the national parliaments have to lead the way for the implementation of national legislation in each member state. The evolution of the release of AI-related legislation is shown in the next figure:

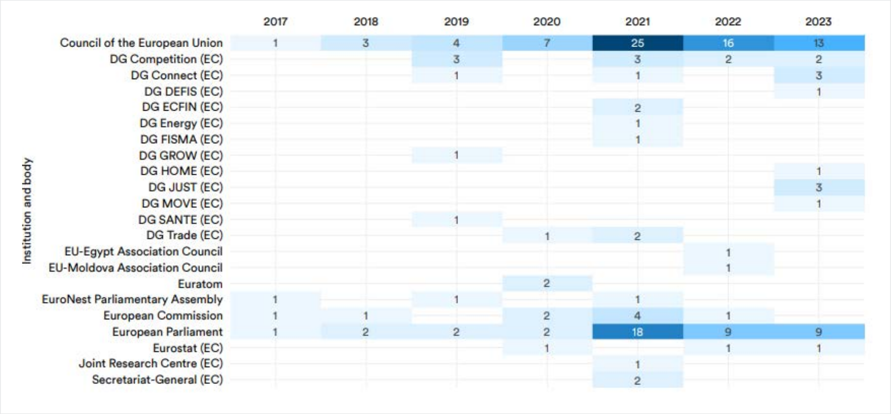
Figure 7: The evolution of AI-related legislation in the European Union



Source: Stanford AI Index, 2024

All the European institutions were involved in releasing AI legislation, and an analysis of each institution's involvement will reflect the community's commitment to implementing AI governance. All the general directorates of the European Commission were analysed in terms of the number of legislative acts generated, as well as the Council of European Union, and the European Parliament. There are some missing elements from this analysis such as the different format of the Council of European Union, or the Competitiveness Council (which also includes the research component). The figure below presents the number of AI-related legislative acts released by the EU's institutions:

Figure 8: The evolution of AI-related legislation in the European Union

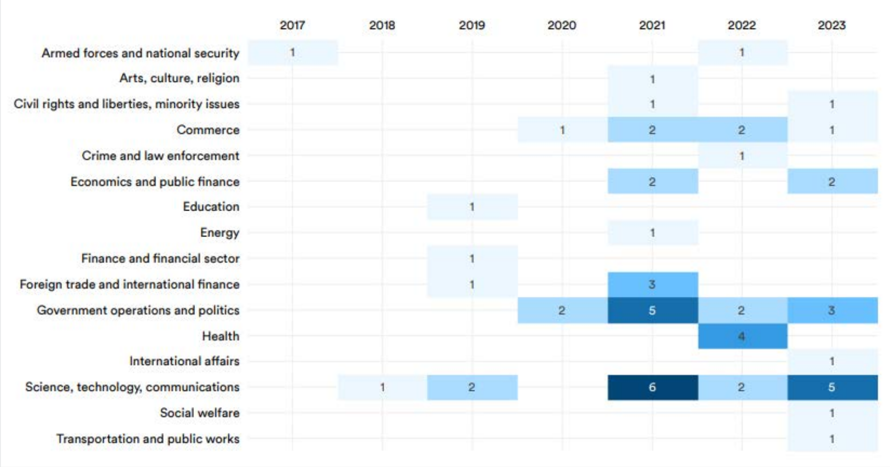


Source: Stanford AI Index, 2024

Figure 8 indicates that the Council of EU adopted the most legislative acts (in 2021 – 13 acts, in 2022 – 16 acts and in 2023 – 13 acts). This demonstrates the agreement with the member states on their different formats. The European Parliament also contributed with its 18 legislative acts and actions in 2021, and 9 in 2022 and 2023 to the growing number of AI-related legislation, expressing the political will of the European institutions. The different DGs also contributed, as departments of the European Commission, to initiating AI-related legislation. The diversity of the DGs shows the involvement of AI in various European policies, such as: Competition Policy (DG Competition – 10 acts in total), Transport Policy (DG Connect – 5 acts in total and DG MOVE – 1 act), Common Defence and Security Policy (DG DEFIS – 1 act), Economic Policy (DG ECFIN – 2 acts), Energy Policy (DG Energy – 1 act), Financial Policy (DG FISMA – 1 act), Internal Market – the four freedoms (DG Grow – 1 act), Migration Policy (DG HOME – 1), Judicial Policy (DG JUST – 3 acts), Health Policy (DG SANTE – 1 act), and Commercial Policy (DG TRADE – 3 acts). The presence of AI in legislation from different sectors of activities and policies reveals its transversality.

To improve the legislation in the field of AI, the identification of the topics approached is highly significant for the European policies that will be implemented, as synthesised in the figure below. Among these topics, science, technology and communications were the most represented – 16 AI-related regulations. The European approach of AI-related legislation cannot be completed without mentioning the AI Innovation package released on January 2024 by the European Commission with the goal of making Europe's supercomputers accessible to innovative European AI start-ups looking to train their AI models (Renda, 2024; European Commission, 2024c). Mirroring the NAIRR Programme from the USA, the European Union has started to create the ecosystem required, yet this will not be enough to boost the European innovation and industrial policy with the use of AI.

Figure 9: The evolution of AI-related legislation in the European Union



Source: Stanford AI Index, 2024

To reach these objectives, many European member states have already commenced the elaboration of their legislative framework in the AI field and this engagement shows the proactive commitment and responsibilities for establishing a solid legislative structure. In the next section, attention is paid to presenting the Romanian initiatives in this area, as part of its efforts to align with the broader European strategy.

3 Romanian AI-Related Legislation

The regulatory dimension of AI in Romania reflects a combination of national initiatives, implementation of EU directives in this field and development of a framework to ensure the deployment of AI is ethical and responsible.

Right from the outset, it should be mentioned that Romania joined the initiative of cooperation in the AI field launched by the Commission in May 2018, together with other member states such as Greece, Cyprus and Croatia

(European Commission, 2018d). Further, on the national level, the launch of the national AI strategy was initiated in 2019 (Maslej et al. ,2024; Stanford AI Index, 2024) before finally being approved by the Romanian Government on 11 July 2024 (HG-SN-IA – 22012024, 2024). Since 2018, all member states have taken a similar approach to introducing their national strategy.

As a member of the EU, Romania adheres to the EU's regulatory framework for AI, which includes:

- The General Data Protection Regulation (GDPR): ensuring AI systems comply with data protection and privacy laws (European Commission, 2016);
- The AI Act: the EU is working on an AI Act to regulate the use of AI based on risk categories. This will require Romania to align its national regulations with EU standards (European Commission, 2018f); and
- Ethics Guidelines for Trustworthy AI: these guidelines outline principles like transparency, accountability, and human oversight, which Romania aims to integrate into its national AI practices (European Commission, 2020b).

The Romanian National Strategy in the field of AI for 2024–2027 (Ministerul Cercetării, Inovării și Digitalizării, 2024), adopted in July 2024, was elaborated following a European project called "Strategic framework for the adoption and use of innovative technologies in public administration 2021–2027 – solutions for the efficiency of the activity", financed under the Operational Programme Administrative Capacity 2014–2020, implemented by the Authority for Digitisation of Romania, in partnership with Technical University of Cluj Napoca. The project's main objective was to align international strategies related to the use of innovative technologies in public administration with the national context, and to develop strategic directions for the period 2021–2027.

The strategy aims to boost AI research and innovation, adopt digital technologies in the economy and society and across various sectors, ensure ethical and responsible AI use, respect for human rights, and promote excellence and trust in AI (Ministerul Cercetării, Inovării și Digitalizării, 2024). It was drafted between July 2021 and February 2023 by a diverse team of experts in technology, research and innovation, digitisation, entrepreneurship, and public policy, and subjected to an extensive multilevel public consultation process. This comprehensive strategy – with its horizontal, cross-sectoral nature – proposes measures for optimally leveraging the existing potential of the entire national AI ecosystem in Romania.

The strategy is organised around six general objectives and several specific objectives, as described below and in a previous study (Ciot, 2023).

Table 4: Objectives of the National Strategy in the field of AI for 2024–2027 for Romania

General objectives	Specific objectives
OG1. Supporting education for RDI and the training of specific AI skills	OS1.1. Increasing training capacity and training level of an AI specialist
	OS1.2. Increasing the level of basic understanding of the population regarding the benefits, use and regulation of AI technologies
OG2. Development and use efficient infrastructure and datasets	OS2.1. Development of AI-specific hardware infrastructure and transparent and fair access to it, to facilitate the processes of R-D-I and production in this field
	OS2.2. Expanding the use of datasets, with application in various sectors of activity
OG3. National Research – Development – Innovation System development in the field of AI	OS3.1. Development of fundamental and applied scientific research specific to the field of AI, as well as on an interdisciplinary level
	OS3.2. Reducing the fragmentation of R&D resources and efforts in AI by conjugating and synchronising them within some centres and national specialised innovation groups connected to the centres and international AI resources
	OS3.3. Supporting and promoting AI innovation
OG4. Transfer insurance technologically through partnerships	OS4.1. Improving the exploitation of research results by developing technological transfer capacities
	OS4.2. The establishment and organisation of a national network of testing spaces and experimentation (TEF) of solutions developed in the field of AI
OG5. Facilitating the adoption of AI across the whole of society	OS5.1. Adoption of AI technology in the public sector
	OS5.2. Adoption and exploitation of AI technologies in sectors in economic priority sectors
OG6. Developing a governance and of a regulation system for AI	OS6.1. Ensuring the governance framework for AI development
	OS6.2. Facilitating AI development through regulation

Sources: Ministerul Cercetării, Inovării și Digitalizării, 2024, Ciot, 2023

The implementation of SN-IA is monitored with specific indicators assigned to each general objective (impact indicators), each specific objective (result indicators), and each measure (immediate achievement indicators). The expected results of implementing the National Strategy in the field of AI in Romania are:

- The development of the research and innovation sector in ICT – focusing on human resources, expertise, and national and international recognition;
- Enhancing the capacity to train and educate AI specialists within the education system;
- The widespread dissemination of basic AI knowledge and skills among the population and businesses;
- The development of AI-specific infrastructures (investment, regulation, datasets);
- The development of the institutional ecosystem with AI expertise, including research centres, companies, and spaces for testing and experimenting with solutions;
- The implementation of AI solutions in the public sector for digital public services and in the private sector for economic competitiveness; and
- The enhanced governance and regulation of AI (Ministerul Cercetării, Inovării și Digitalizării, 2024).

As may be seen, the educational dimension is pivotal in the strategy, at different levels of instruction, with large categories of persons, from children to adults; also, the development of the infrastructure to create a proper ecosystem for the development of AI in order to be used by public services, and the private sector, to improve future governance and competitiveness.

The Strategy includes several policy recommendations and conclusions, envisages its future implementation as well as its political and social dimensions:

- The Strategy is to be fully implemented within the agreed timeframe, assuring that Romania's progress in the field is synchronised with developments on the European level and addresses national specifics;
- Stakeholders like the academic sector, public administration and business environment will collaboratively and sustainably participate in implementing and monitoring measures, as well as in applying regulations;
- Strategy implementation will maintain flexibility to accommodate operational adaptations based on field dynamics, unforeseen developments,

technological advancements, regulatory updates, societal technology adoption rates, and effective progress in launching and implementing AI projects and programmes;

- The monitoring and evaluation function in implementation of SN-IA (the Strategy) will be actively supported and continuously strengthened, serving as a crucial element in overseeing overall progress, providing timely information to decision-makers, and providing for successful implementation; and
- Effective leadership is essential for aligning goals and approaches to implementation, and for steadfastly upholding the application of the proposed measures (Ministerul Cercetării, Inovării și Digitalizării, 2024).

These recommendations reveal the opportunities Romania is using to contribute to this sector's progress on the European level by involving Romanian expertise and knowledge to achieve a better position at the global forefront of AI innovation.

The impact of AI technologies is also important in the private sector where Romania, according to Pislariu (2024, apud Eurostat, 2024), was in the lowest place in the classification on the European level with only 1.5% of companies using these technologies compared with the European average of 8% or the first places occupied by Denmark and Finland with 15% of companies using AI. Pislariu (2024) mentions the Forbes Report which for 2027 projected the AI market would grow to USD 407 billion by 2027 and draws attention to the fact that there are some structural and cultural issues regarding the existence of a gap between Romania and other member states. Proper legislation, education and infrastructure is a must for Romania for it to reduce these gaps in AI development.

Adopting a national artificial intelligence (AI) strategy is critical for several aspects of any state. This not only establishes a coherent framework and strategic directions for the use and development of AI, but also brings numerous practical and strategic benefits. First, a national strategy for AI helps strengthen the national technological and innovative capabilities of a country. By identifying and promoting local AI expertise, the strategy can boost investment in research and development, attract talent and promote collaboration between the public, private and academic sectors. Second, adopting such a strategy can enhance a nation's economic competitiveness. The use of AI in various economic sectors, such as health, transport, agriculture or public administration, can increase operational efficiency, reduce costs and improve services for citizens and companies. In addition, a national AI strategy can ensure that the development and use of artificial intelligence technologies respect ethical principles and fundamental values such as personal data protection, transparency and non-discrimination. This might also include establishing clear legislative and regulatory frameworks to manage the risks associated with AI use. Finally, adopting a national AI strategy bolsters a country's position within the international and European community, facilitating cross-border collaborations in research and innovation, and helping to shape global AI directions.

The development of a national strategy in the field of AI not only consolidates the progress already achieved through existing projects, but also guides and supports future initiatives in a coherent and effective way, assuring that the technological development of the country is well managed and brings significant benefits to society and the economy.

4 Romanian AI-Related Projects – Good Practice Models

The importance of AI in Romania requires a proper legislative framework for its application. Several initiatives and projects have been developed in sectors like education, medicine, agriculture, transport, research and innovation, public administration, and the private sector.

On an educational level, e-learning platforms and virtual assistants are the most recent tendences to have appeared, yet consolidated experiences have also already been had, especially with the big universities. Adservio is a e-learning platform recognised by the Ministry of Education targeting schools and kindergarten education. School Intuitext is a platform that creates exercises and tests for pupils based on AI algorithms (Școala Intuitext, 2024).

University Babes-Bolyai developed a BA programme for AI starting in September 2023, which raised considerable interest among future students and entailing a competition with 8 persons vying for a place (NewsUBB, 2023). As a member of a consortium comprising 17 research institutes, industrial partners, and leading European universities, since 2023 BBU is implementing a large-scale European project in the field of AI. Known as Artificial Intelligence for Connected Industries (AI4CI), the project aims to pioneer innovative interdisciplinary programmes. Its goal is to address challenges and capitalise on opportunities arising from the use of AI in connected systems and equipment. AI4CI focuses on developing high-impact vertical applications for automated industrial systems, IoT (Industrial Internet of Things) systems, and connected autonomous vehicles. Aligned with the European AI strategy, the mission of AI4CI is pivotal for advancing AI development within industrial contexts, crucial for the EU's ambition to lead in AI. Over the next 4 years, the AI4CI project is to be implemented with a total budget of EUR 10 million.

The programme will provide elite training for future AI specialists and meet the needs of businesses and entrepreneurs while maximising the potential of research centres and digital innovation hubs.

Technical University from Cluj-Napoca developed a project called Automated Urban Parking and Driving – UP-Drive in partnership with Volkswagen, IBM, Technical University from Prague, and Federal Polytechnic of Zurich. The project ended in 2019 with the presentation of a fully autonomous VW car that has passed tests conducted in controlled conditions (G4Media, 2020).

University Politehnica of Bucharest has developed a MA programme on AI and a project called Artificial Intelligence and Multi-Agent Systems Laboratory (University Politehnca of Bucharest, 2024), which generated several other projects such as ENFIELD – European Lighthouse to Manifest Trustworthy and Green AI, Virtual Patient – An AI-based System for Training in Cardiovascular Diseases Diagnostic and Treatment, Bridging the Early Diagnosis and Treatment Gaps of Brain Diseases, MERITT – Maintaining Active and Healthy Living Through Serious Games and Artificial Intelligence, CNCC – Creation, Operationalisation and Development of the National Centre of Competence in the Field of Cancer, CASHMERE – Context-Aware Search and Discovery in Hypermedia-Driven Multi-Agent Environments, AI Folk – Resource Management in Distributed AI, and DigiTwin4CIUE – Digital Twins for Complex Infrastructures and Urban Ecosystems. Many research publications were created during this project, as mentioned at the Laboratory's website: https://aimas.cs.pub.ro/publications/.

In the medical field, there is the National System of Telemedicine (Ministerul Sănătății, 2022), implemented to facilitate remote medical consultations. This system employs AI algorithms to analyse patient data and provide preliminary diagnoses, and Medical Imaging Analysis, through projects that use AI to analyse medical images (e.g. X-rays, CT scans) to detect diseases such as cancer early. MediNav is a project that develops a telemedicine platform incorporating AI to analyse patient data and offer initial diagnoses, designed to enable remote medical consultations, particularly in rural areas (MediNav, 2024). Further, SkinVision is a project employing AI to analyse skin images and identify early signs of skin cancer (SkinVision, 2024). There are several good initiatives and projects, such as MedicAI – Start-up în domeniul medical/Start-up in the medical field, which is a Romanian–American platform that facilitates connections among doctors for collaborative analysis of medical imaging. It also supports collaboration between doctors and patients (Start-Up, 2020). Another project is Xvision – Artificial Intelligence for Smarter Healthcare, which is a startup that utilises AI to provide information on medical imaging, identify issues, and enhance the efficiency of doctors' work (X-vision, 2024). Also worth mentioning is the project KEY4VISION from Cluj-Napoca, a start-up that is developing an enhanced prototype of a 3D periodontal ultrasound scanner, incorporating neural networks to automate the processing of acquired data from ultrasonographic images (KEY4VISION, 2024). A project in telemedicine which uses AI is CardioMedive, a Romanian company developing one of the most advanced patient monitoring systems, embodied in a discrete patch that continuously measures a comprehensive range of vital and hemodynamic signs. Perhaps the most important initiative in this sector is Project Deep Health financed by the Horizon Programme, with 22 partners, including hospital and medicine universities, with a total budget of EUR 14 million, aiming to meet genuine healthcare sector needs and simplify the daily tasks of medical staff and expert users in terms of image processing, and using and training predictive models, all within a unified toolset. From Romania, Prof Dr Theodor Burghele from the Clinic Hospital from Bucharest was involved. The project was implemented between 2019 and 2021.

In the environment sector, EcoAssist utilises AI to monitor and safeguard the environment, which includes analysing data on air and water quality and detecting climate change.

Agriculture is another area where AI is used successfully in Romania. Two policy directions are in place: (a) precision agriculture – utilising drones and sensors to collect field data, which is then analysed by AI algorithms to optimise irrigation, fertilisation, and plant protection; and (b) crop health monitoring – initiatives employing AI to monitor crop health and detect early signs of diseases or pests. Agro AI relies on drones and sensors to collect data from agricultural fields, which is then analysed by AI algorithms to optimise irrigation, fertilisation, and plant protection. FarmVisionAI is a project dedicated to monitoring crop health and identifying early signs of disease or pests via the analysis of images captured by drones. Several start-ups are revolutionising the principles and mechanisms. AgroCity is a Romanian start-up that integrates cutting-edge technologies to oversee farm operation suppliers, and land parcels. It facilitates real-time monitoring to minimise losses and boost productivity (AgroCity, 2024). Agricloud is a Romanian platform providing analytical data on grapevine crops. Utilising Internet of Things (IoT) technology, the platform monitors crops while improving planning, procurement, and inventory control. It optimises production, coordinates irrigation sensors, and manages controls effectively (Agricloud, 2024)

In the field of public services, AI is used as: (a) chat boxes for public services – implementing chatbots to assist citizens with information and support for interacting with public services, including the paying of taxes and requesting of official documents; and (b) public data analytics – initiatives utilising AI to analyse large volumes of public data to support decision-making and optimise public services. Ana is a chat box created by Bucharest City Hall to offer information and aid to citizens regarding various administrative matters. ConsulAI is a pilot project initiated by the Ministry of Justice to deliver legal support and information via a virtual assistant. In Romania, many projects are already implemented: the E-governance project, which is the National Electronic System that computerises the interaction between the citizen/company and the public administration and makes forms used in the relationship with the administration available to the citizen, facilitating the obtaining of various documents (Autoritatea pentru Digitalizarea României, 2024). SEAP is an electronic platform that ensures the transparency of the public procurement process and procedures. The platform *www.ghiseul.ro* allows taxpayers to view their current payment obligations and make online payments using a credit card, covering local taxes, duties, and outstanding fines either partially or in full. The platform aici.gov.ro serves as an intermediary for registering documents intended for public institutions that lack their own online registration system. Apiary Book is a beehive management solution designed to reduce colony losses caused by diseases, pests and pesticides, enabling informed decisions to add to productivity and maximise profitability (ROTSA, 2024). Finally, Ogor is a satellite monitoring service dedicated to farmers that offers a comprehensive record of land conditions dating back to 2017 (Ogor, 2024).

In the transport and mobility sector, there are initiatives such as: Traffic Management Systems, which employs AI to monitor and manage urban traffic in real time, reducing congestion and optimising traffic flow, and autonomous vehicles – pilot projects testing autonomous vehicles on public roads, utilising AI for navigation and safety.

In the research and innovation field, there are several important collaborations between universities and research institutes seeking to develop new AI-based technologies. Examples include the Polytechnic University of Bucharest and the Romanian Academy's Institute for Artificial Intelligence Research and hackathons and competitions, hosting events that foster AI innovation where participants create innovative solutions within a short timeframe.

For the private sector, the tendencies to implement AI technologies can be grouped in the following directions: business process automation (for example, by implementing AI-based robotic process automation/RPA solutions to streamline repetitive tasks and enhance operational efficiency), and predictive analytics, to leverage AI to analyse sales and marketing data, for predicting market trends and consumer behaviour. One good example is the company UiPath, a globally recognised Romanian enterprise specialising in robotic process automation (RPA), leveraging AI to automate repetitive business processes, thereby boosting efficiency and cutting operational costs.

These initiatives illustrate Romania's dedication to adopting and integrating AI technologies across diverse sectors, fostering the development of a digital economy and enhancing citizens' quality of life.

5 Conclusion

The adoption of a national strategy in the AI field is essential for the sustainable and competitive development of a country in the digital age, bringing benefits on multiple levels and directing technological progress in advantageous directions for society and the economy. Romania's approach to AI regulation emphasises the importance of establishing a robust legal and ethical framework.

The regulatory efforts in Romania seek a balance between fostering innovation and protecting the interests of society. With a supportive ecosystem for the development of AI in Romania in place, economic growth will be stimulated and competitiveness will follow.

Making sure that the AI system adheres to the ethical principals is a cornerstone of the Romanian legislation to succeed as the first national AI strategy. This includes promoting accountability, safeguarding privacy rights, and mitigating biases to build trust among users and stakeholders.

International collaboration is vital for developing a suitable AI regulation. By aligning with the EU's directives, Romania intends to align with standards and facilitate global exchanges. Adaptability and flexibility will characterise the harmonisation of regulation with future requirements.

The importance of public engagement and education in AI regulation is recognised on the national level, encapsulating a proactive stance. Effective AI regulation requires continuing monitoring and evaluation of the impact of regulatory measures, and Romania is committed to make the necessary adjustments so as to stay responsive to technological advancements and the needs of society.

REFERENCES

- Adservio. (2024). Platforma Nr. 1 de management școlar cu Modul Financiar integrat Potrivită pentru școlile și grădinițele publice și private. Retrieved 12 July 2024 from https://www.adservio.ro/ro/ .
- AgroCity. (2024). Simplifică gestionarea fermei tale într-o singură aplicație Programul tău pentru agricultură digital. Retrieved 12 July 2024 from https://agrocity.eu/ .
- Agricloud. 2024. Agricultură de precizie=Rentabilitate crescută. Retrieved 12 July 2024 from https://agricloud.ro/ .
- Autoritatea pentru Digitalizarea României. (2024). Platforme și servcii. Retrieved 12 July 2024 from https://www.adr.gov.ro/ .
- Boyer, A., Cortez, E.K., DeCrescenzo, R., Dever, C., Freeman Engstrom, D., Fattorini, L., de Guzman, P., Hassan, M., He-Li Hellman, E.D., Ho, D., Hsu, J., Jonga, S., Kosoglu, R., Lemley, M., Betts Lotufo, J., Maslej, N., Meinhardt, C., Nyarko, J., Park, J., Parli, V., Perrault, R., Rome, A., Santarlasci, L., Smedley, S., Wald, R., Williamson, E., Zhang, D. (2024). Policy and Governance. In Maslej, N., et al. (2024). The AI Index 2024 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford, University, Stanford, CA, April 2024, pp. 366-411. Retrieved 12 July 2024 from https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf .
- Ciot, M.G. (2023). Projections of AI's applications and regulations in Romania. In Dr. Aleksander Aristovnik, Dr. Dejan Ravšelj and Dr. Nina Tomaževič (eds.), Artificial Intelligence for human-centric society: The future is here, ISBN: 978-2-39067-063-6 9782390670636, Bruxelles, 2023, pp. 260-291.
- Committee of Regions (2021). European Approach to Artificial Intelligence - Artificial Intelligence Act (revised opinion). Retrieved 1 July 2024 from https://cor.europa.eu/EN/our-work/Pages/OpinionTimeline.aspx?opId=CDR-2682-2021 .
- Council of European Union (2022). Artificial Intelligence Act: Council calls for promoting safe AI that respects fundamental rights. Retrieved 1 July 2024 from https://www.consilium.europa.eu/en/press/press-releases/2022/12/06/artificial-intelligence-act-council-calls-for-promoting-safe-ai-that-respects-fundamental-rights/ .
- Council of European Union (2021). Council of the EU: SI Presidency compromise text on the AI Act. Retrieved 1 July 2024 from https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf .
- Euroactiv. (2024). EU countries give crucial nod to first-of-a-kind Artificial Intelligence law. Retrieved 1 July 2024 from https://www.euractiv.com/section/artificial-intelligence/news/eu-countries-give-crucial-nod-to-first-of-a-kind-artificial-intelligence-law/ .
- European Central Bank. (2021). Opinion of the European Central Bank of 29 December 2021 on a proposal for a regulation laying down harmonised rules on artificial intelligence. (CON/2021/40). (2022/C 115/05). Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021AB0040

- European Commission. (2024a). European approach to artificial intelligence. Retrieved 3 July 2024 from https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence .
- European Commission. (2024b). Setting up of AI factories now possible after EuroHPC Regulation amendment. Retrieved 3 July 2024 from https://digital-strategy.ec.europa.eu/en/news/setting-ai-factories-now-possible-after-eurohpc-regulation-amendment .
- European Commission. (2024 c). AI innovation package to support Artificial Intelligence startups and SMEs. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/news/commission-launches-ai-innovation-package-support-artificial-intelligence-startups-and-smes .
- European Commission. (2024 d). European AI Office. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/policies/ai-office .
- European Commission. (2023a). Commission welcomes political agreement on Artificial Intelligence Act*. Retrieved 1 July 2024 from https://ec.europa.eu/commission/presscorner/detail/en/ip_23_6473 .
- European Commission. (2022a). Launch event for the Spanish Regulatory Sandbox on Artificial Intelligence. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/events/launch-event-spanish-regulatory-sandbox-artificial-intelligence .
- European Commission. (2022b). Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). COM/2022/496 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52022PC0496 .
- European Commission. (2021 a). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European approach to Artificial Intelligence. COM/2021/205 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM%3A2021%3A205%3AFIN .
- European Commission. (2021 b). Proposal for a Regulation of the European Parliament and of the Council laying down harmonized rules on artificial intelligence (artificial intelligence act) and amending certain union legislative acts. COM/2021/206 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206 .
- European Commission. (2021 c). Proposal for a Regulation of the European Parliament and of the CounciL on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council, and repealing Council Directive 87/357/EEC and Directive 2001/95/EC of the European Parliament and of the Council. COM/2021/346 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0346&qid=1628522210573 .
- European Commission. (2021d). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions 2030 Digital Compass: the European way for the Digital Decade. COM/2021/118 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52021DC0118 .
- European Commission. (2021e). Horizon Europe. Retrieved 1 July 2024 from https://research-and-innovation.ec.europa.eu/funding/funding-opportunities/funding-programmes-and-open-calls/horizon-europe_en .
- European Commission. (2020a). White Paper on Artificial Intelligence: a European approach to excellence and trust. COM(2020) 65 final. Retrieved 1 July 2024 from https://commission.europa.eu/document/download/d2ec4039-c5be-423a-81ef-b9e44e79825b_en?filename=commission-white-paper-artificial-intelligence-feb2020_en.pdf

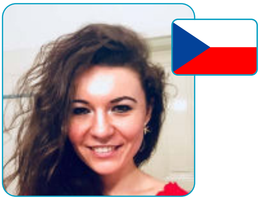
- European Commission. (2020b). Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment
- European Commission. (2020c). Second European AI Alliance Assembly. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/events/second-european-ai-alliance-assembly .
- European Commission. (2019a). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions - Building Trust in Human Centric Artificial Intelligence [COM(2019)168]. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX%3A52019DC0168
- European Commission. (2019b). The first European AI Alliance Assembly. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/node/1851/printable/pdf
- European Commission. (2018a). Artificial intelligence: Commission kicks off work on marrying cutting-edge technology and ethical standards. Retrieved 3 July 2024 from https://ec.europa.eu/commission/presscorner/detail/en/IP_18_1381 .
- European Commission. (2018b). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Artificial Intelligence for Europe, COM/2018/237 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018DC0237 .
- European Commission. (2018c). Commission staff working document. Liability for emerging digital technologies. Retrieved 1 July 2024 from file:///C:/Users/Zsolt/Downloads/swd2018137_DB3D7786-BAF7-E688-9ECC809B21BF57FA_51633.pdf .
- European Commission. (2018d). Declaration of cooperation on Artificial Intelligence (AI). Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/node/239/printable/pdf .
- European Commission. (2018e). The European AI Alliance. Retrieved 1 July 2024 from https://digital-strategy.ec.europa.eu/en/policies/european-ai-alliance .
- European Commission. (2018f). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions. Coordinated Plan on Artificial Intelligence, COM/2018/795 final. Retrieved 1 July 2024 from https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795 .
- European Commission. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). Retrieved 12 July from https://eur-lex.europa.eu/eli/reg/2016/679/oj .
- European Economic and Social Committee (2021). Regulation on artificial intelligence. Retrieved 1 July 2024 from https://www.eesc.europa.eu/en/our-work/opinions-information-reports/opinions/regulation-artificial-intelligence .
- European Parliament. (2023). MEPs ready to negotiate first-ever rules for safe and transparent AI. Retrieved 1 July 2024 from https://www.europarl.europa.eu/news/en/press-room/20230609IPR96212/meps-ready-to-negotiate-first-ever-rules-for-safe-and-transparent-ai .
- G4Media. (2020). VIDEO: IT-iști clujeni de la Universitatea Tehnică au dezvoltat programul care percepe realitatea pentru Volkswagenul fără șofer. Retrieved 12 July 2024 from https://www.g4media.ro/video-it-isti-clujeni-de-la-universitatea-tehnica-au-dezvoltat-programul-care-percepe-realitatea-pentru-volkswagenul-fara-sofer.html .

- KEY4VISION. (2024). #D ultrasound imagining technology. Retrieved 12 July 2024 from https://www.key4vision.com/ .
- Ministerul Cercetării, Inovării și Digitalizării (2024). Stratega Națională în domeniul Inteligenței Artificiale 2024-2027. Retrieved 12 July 2024 from https://www.mcid.gov.ro/wp-content/uploads/2024/01/Strategie-Inteligenta-Artificiala-22012024.pdf .
- Ortega Klein, A. (2020). Digital decolonisation: The EU's new ideas on data and artificial intelligence. Retrieved 28 June 2024 from https://ecfr.eu/article/commentary_digital_decolonisation_the_eus_new_ideas_on_data_and_artificial/ .
- Renda, A. (2024). Towards a European large-scale initiative on artificial intelligence ceps in-depth analysis. What are the options? Retrieved 12 July 2024 from https://cdn.ceps.eu/wp-content/uploads/2024/07/2024-10_ID-A_Towards-a-European-large-scale-initiative-on-Artificial-Intelligence.pdf .
- Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J.C., Shoham, Y., Wald, R. and Clark, J. (2024). The AI Index 2024 Annual Report, AI Index Steering Committee, Institute for Human-Centered AI, Stanford, University, Stanford, CA, April 2024. Retrieved 12 July 2024 from https://aiindex.stanford.edu/wp-content/uploads/2024/05/HAI_AI-Index-Report-2024.pdf .
- MediNav (2024). Redăm timpul medicului pentru pacienți Asistentul tău medical care învață și reduce timpul de documentare!. Retrieved 12 July 2024 from https://medinav.eu/
- Ministerul Sănătății. (2022). Normele de aplicare a serviciilor de telemedicină, aprobate. Retrieved 12 July 2024 from https://ms.ro/ro/centrul-de-presa/normele-de-aplicare-a-serviciilor-de-telemedicin%C4%83-aprobate/ .
- NewsUBB. 2023. Prima generație de studenți la licență în inteligența artificială din România își începe studiile la UBB. Retrieved 12 July 2024 from https://news.ubbcluj.ro/prima-generatie-de-studenti-la-licenta-in-inteligenta-artificiala-din-romania-isi-incepe-studiile-la-ubb/ .
- OGOR. (2024). Controlezi cultura și efectele lucrărilor Agricole Aplicație de analiză agronomică bazată pe imagini din satelit. Retrieved 12 July 2024 from https://ogor.ro/ .
- Pilariu, A. (2024). The digital evolution of Romania: Artificial Intelligence, a key piece in the puzzle of economic success. Retrieved 12 July 2024 from https://connectionsconsult.ro/the-digital-evolution-of-romania-artificial-intelligence-a-key-piece-in-the-puzzle-of-economic-success/ .
- ROTSA. 2024. Apiary Book Tech Startups to watch in the Romanian Ecosystem. Retrieved 12 July 2024 from https://rotsa.ro/news/apiary-book-business-app-beekeeping/ .
- Secretariatul General al Guvernului. (2024) – HOTĂRÂRE privind aprobarea Strategiei Naționale pentru Inteligența Artificială. Retrieved 12 July 2024 from https://www.mcid.gov.ro/wp-content/uploads/2024/01/HG-SN-IA-22012024.pdf .
- SkinVision.(2024). Skin Cancer Melanoma Tracking App Smart about skin health Retrieved 12 July 2024 https://www.skinvision.com/ .
- Start-Up. (2020). Startup-ul Medicai, tratamentul prin inteligență artificială pentru medicină. Retrieved 12 July 2024 from https://start-up.ro/startup-ul-medicai-tratamentul-prin-inteligenta-artificala-pentru-medicina/ .
- Școala Intuitext. (2024). PREDARE. ÎNVĂȚARE. EVALUARE. Asistent educațional digital pentru clasele primare. Retrieved 12 July 2024 from https://www.scoalaintuitext.ro/ .
- Universitatea Bebș-Bolyai. (2023). Proiect european de anvergură în domeniul inteligenței artificiale, implementat de UBB. Retrieved 12 July 2024 from https://www.cs.ubbcluj.ro/proiect-european-de-anvergura-in-domeniul-inteligentei-artificiale-implementat-de-ubb/ .

- University Politehnca of Bucharest. (2024). AI-MAS - Artificial Intelligence and Multi-Agent Systems Laboratory. Retrieved 12 July 2024 from https://aimas.cs.pub.ro/ .
- University Politehnca of Bucharest. (2024). Master of Science in Artificial Intelligence Programme. Retrieved 12 July 2024 from https://aimas.cs.pub.ro/master_ai/ .
- US National Science Foundation. (2024). National Artificial Intelligence Research Resource Pilot. Retrieved 1 July 2024 from https://new.nsf.gov/focus-areas/artificial-intelligence/nairr#:~:text=The%20NAIRR%20is%20a%20concept,to%20participate%20in%20AI%20 research.
- X-vision (2024). Artificial Intelligence for Smarter Healthcare. Retrieved 12 July 2024 from https://xvision.app/.

Chapter 8

We Do Not Want Big Brother in the EU



Šárka Shoup

1 Introduction

The European Union is the first in the world to implement comprehensive rules for the use of artificial intelligence (AI). While negotiating the AI Act, representatives agreed that clear rules would provide tech companies with predictability and citizens a sense of security while utilising AI. According to the representatives, regulation can ultimately benefit innovation. The AI Act was unanimously approved by member states at the beginning of February 2024, before coming into effect on 1 August 2024.

The biggest topic of dispute during the negotiations was Chat GPT. Some member states were hesitant to introduce regulations for these technologies as they did not want to decelerate their development and innovations. However, they worked together to find an agreement. It is impossible to say that everyone is 100% satisfied, although there is consensus that a balance must be found between the development of innovations and the preservation of security. Did the EU representatives succeed? Is life easier without 'rules' or do rules instil greater confidence in us?

By establishing regulations and rules for artificial intelligence in the European Union market, companies will know in advance what is (not) allowed. This will ensure that companies can proactively work with this information while creating their budget. The technologies that are planned to be banned are primarily surveillance systems that are in common use, for example, in China. The purpose of setting the rules is to instil trust in AI among the public and create a stronger desire concerning usage. Hence, the sharing of information will increase.

This chapter asks in which areas can AI really benefit us and to what extent we can let it penetrate our lives. It also identifies where we are already reaching the edge in sharing personal information and how certain regulations are needed by both the EU and the nation state.

2 About Artificial Intelligence and Its Threats

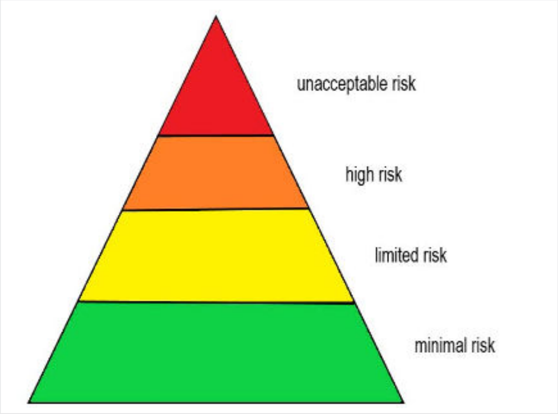
Lately, we have been hearing from all sides about AI and its threats: it will take our jobs, manipulate elections, and spy on us at every step. These are the concerns. Therefore, on 1 August 2024, the European introduced the world's first-ever legislation to establish rules for the use of artificial intelligence. What will it protect us from? And will the new regulations not hinder the development of innovations? Will the EU shoot itself in the foot doing this?

2.1 Goal of the AI Act

The goal of the AI Act is to develop and safely use artificial intelligence in the EU market by ensuring legal certainty and strengthening the fundamental rights, safety and health of European citizens. With clear rules about what can(not) be done with AI in the EU market, companies will be able to plan accordingly while developing their products. Easy-to-understand and specific rules across the EU will reduce costs for companies in terms of dealing with the regulatory environment. When clear rules are in place, people will be able to trust how AI is handled more, increasing their interest and willingness to buy and use AI products, thus simplifying their lives.

According to the AI Act, systems are classified in four risk categories. All AI systems that fall under unacceptable risk will be banned, with this mainly including manipulating vulnerabilities, social scoring, assessing criminality based solely on profiling, or real-time biometric identification in a public space. Significant changes are coming especially for high-risk AI systems, which include those in core infrastructure, education, and law enforcement; they will be subject to more stringent regulation by the AI Act. This is particularly the case in the areas of data governance, human monitoring or openness requirements. For the time being, AI systems such as chatbots and deepfakes remain in the limited risk category. The legislation will also actively encourage the development of low-risk (minimal) AI applications, which should prevent innovation from stagnating, but also protect fundamental rights and strengthen the very ethicality of AI. Systems will necessarily be designed and used in a way that eliminates any mass surveillance or other activities that could potentially violate the rights of individuals.

Figure 1: Levels of risk for AI systems



Source: European Commission a., (2024)

The economic aspect of the AI Act is also important since it focuses on supporting small and medium-sized enterprises (SMEs) and start-ups. The EU wishes to ensure that these businesses can compete with larger technology companies on a level playing field by providing clear guidelines and support structures. Whether we manage to establish such rules that effectively protect our rights without hindering innovation and support development, as planned, is, of course, up for question and depends on one's perspective. The final text was unanimously supported by ambassadors from all 27 EU member countries, although it was a compromise and, like any compromise, is not perfect. The proposal had been fine-tuned for several years, with additional discussions over 2 years following its release. This effort involved extensive work by many experts participating in the entire process. The proposal is also supplemented by a coordinated AI plan issued by the European Commission. In addition, EU member states have adopted AI strategies and proposed several investment and support mechanisms to promote the development and use of artificial intelligence in the EU market.

2.2 Attention to the AI Act

The AI Act was highly anticipated, not only from a political perspective involving almost all ministers but also from lobbyists and companies eagerly awaiting the outcome. There were lively debates on both member state and EU levels, with frequent consultations with stakeholders, each having their own opinion. As early as 2023, some leading European companies, including Siemens, Renault and Meta, did not agree to the proposed AI Act, referring to the fact it could harm the bloc’s competitiveness and stimulate an outflow of investment. Policymakers frequently reached out to companies to understand the practical impacts of the rules being set. Hence, we can say that a common vision emerged with respect to how the final text should look, even though not everyone is entirely satisfied.

2.3 The AI Act Is Unique and the First in the World, or Not?

Commissioner Tiery Breton, after the agreement on AI was concluded, shared a piece of paper that read: “The whole world has artificial intelligence and Europe has regulation”. This picture illustrates that we, as Europe, are truly the first in the world to have clear, binding rules on how we should use AI on the system and model levels. This is groundbreaking, and basically nobody else in the world has anything like it. On the other hand, there are a number of different initiatives on the level of individual states and worldwide that address this matter, but do not do so with legislation. For example, in the United States, there is the recently issued executive order from President Biden which sets the standards, or let us say the rules, for what AI should look like, yet it does not have the element of penalties and corrective tools like we have in Europe. Globally, there are a number of initiatives, such as the G7 association, which includes the major economies of the world. They have the ‘Hiroshima process’, which is a discussion about how to ensure that the development of AI is safe and trustworthy. In other words, they are talking about the same things as us, except that we are a little further along in Europe. The question is whether Europe will be a pioneer on this issue, with other parts of the world following suit. This is the hope of everyone involved in the AI Act. It is vital that discussions take place on an international level, particularly between the USA, the EU, and the G7 countries. This is absolutely crucial because AI is a technology that crosses borders, making it essential that the underlying approaches are internationally aligned. We do not want one member to act in a completely different method contrary to another due to the implications the technology itself holds, and further for the state independently.

The question is whether there is an advantage to implementing these rules sooner rather than later. We may not know the answer until sometime in the future. Some argue that if rules are not established initially, development might be more efficient. It could be more unregulated, but ultimately everything may still work out. In Europe, however, the approach is somewhat different. Here, the emphasis is on anticipating the major issues, which was the aim of adopting the AI Act. Whether this will prove successful remains to be seen. Time will tell, but considering how the discussions around the GDPR went, which focused on protecting personal data, initial strong emotions and negative reactions eventually subsided, then revealing that implementation was not as problematic as feared. In fact, American companies adopted the European rules, now enforcing them outright. If the AI Act follows a similar trajectory, finding a balance between innovation and the protection of rights, it could indeed be the success we are all hoping for.

We can say that the EU is clearly setting a regulatory standard that other parts of the world are trying to follow to some extent. It is a fast-moving field where everyone wants to move ahead in innovation. A more flexible regulatory framework allows this to be more efficient than when the rules are too prohibitive. Is Europe achieving the right balance in this regard? As already noted, it might be too early to judge. It nevertheless must be recalled that the AI Act is in the early day of being enforced. Other regions have prioritised innovation, but elsewhere they are also increasingly focusing on some more pressing issues. The EU understands that it is necessary to set such rules so that even local players can continue to innovate quickly.

A hypothetical question might be whether the EU would behave differently if it had a prominent AI company of its own, of which there are many in the USA. While the USA supports its companies and gives them more freedom, the EU is seen as regulating them more. However, encouraging the emergence of other AI champions in Europe is very important. There are many layers within AI technology, not only the applications and services themselves, but also the infrastructure that keeps them running, which is no less crucial. There are already examples of innovative European companies developing basic AI models. For example, Microsoft recently invested in the French start-up Mistral. Yet, AI requires partnership, and there are only a few companies in the world that can do it alone. The more variety there is in the market and the more players there are, the more choice there will be, which is always good for customers and the economy. This means it is necessary to support the emergence of other important players in Europe, and the EU should not be opposed to it. However, it is necessary to listen to what particular emerging companies need, where they encounter obstacles in their growth, and how they can help each other.

It would certainly benefit from more dialogue with those on the cutting edge of development. In Brussels, people are sometimes too far away from where innovation occurs. There is also a need to communicate with other capital cities across Europe, where you can often see great enthusiasm with AI. There is a lot of talent in Europe, and losing them for any reason would be disappointing to everyone involved, and who could be involved. The European Union should, therefore, have more discussions with leading companies to understand better which obstacles they face and what they can do to ensure their products are used, and grow globally.

The impulse that we have a responsibility and are further required to closely monitor multinational corporations and global concerns is a valid and constructive intention. But something else is underlying – the feeling that regulatory protection is as guiding as in the past, especially if it is starting to come at the expense of practicality and our future. The vast majority of companies that set the pace come from the USA or Asia. Even Tesla does not want to operate its autonomous driving system in Europe due to its strict regulations. If that were not enough, after years of relative calm, Europe has entered into a huge dispute with Microsoft, another technological pioneer. At the same time, it does not have to be all about ones and zeros. One of the most cited Czech scientists, the biologist Jiří Friml, who recently received the prestigious Wittgenstein Prize in Austria, has repeatedly pointed out that the European rejection of genetically modified crops is depriving us of the chance to participate in changes that may fundamentally affect both our ecology and economy. The world we live in is developing extremely dynamically. Regulations that react to the surrounding events, on the other hand, are in principle, delayed, so that one who focuses too much on them remains in the past and loses contact with where our zeitgeist, the spirit of the times, is going. There is no question that Europe's past and its present are celebrated, but we should urgently start doing something to make our future also a cause for optimism.

It could be said that the cradle of AI is the USA. So why has regulation with a global ambition not come from there? Americans generally take a significantly liberal approach to technological regulation. They believe that regulation should eventually follow ex-post. Therefore, if new technology does not conflict directly with existing regulations, the authorities must not interfere with its development and expansion. And in the meantime, they will check with their own mechanisms whether or not it will have any significant negative effects. If so, the responsibility lies more with its users. There are lawsuits and subsequent settlements with injured parties and the adoption of remedial solutions. In other words, in the USA they do not primarily regulate preventively, with one of the intentions being to not inhibit innovation.

Of course, even in the USA there is the legal anchoring of AI, yet the legislation there primarily addresses the areas in which why AI will be funded from public budgets and in which circumstances it can be used by government authorities. Specifically, use within government bodies requires the fulfilment of advance preparation standards. However, the private sector is not subject to major regulations.

The AI Act as a regulatory tool is somewhat of an obstacle to the development of AI technologies in Europe, but the question is whether this is good or bad. The answer depends on the moral values of the person we are asking. On one hand, there is the view that the European approach will strangle both the public and the private sectors, which drive progress. On the other hand, there is the fact that the private sphere has a tendency towards limited self-regulation, and most of the time, when some regulation does come out of it, it is more to the benefit of corporations than individuals. Or it is set up so that it does not significantly limit the achievement of the corporation's goals.

A typical example of certain flexibility on the topic of ethical rules is the collection of data on which globally used AI models learn. When considering this, we must not forget the different legal systems. In the USA, it is possible to file a class action lawsuit, which allows many potentially injured persons to be represented in a single action, even without their direct participation. This is impossible in the European sector/theatre, and potential lawsuits or other legal actions are not a sufficient deterrent for companies. Another argument is that the EU, on one hand, hinders development. Still, on the other hand, it increases legal certainty and gives investors clear boundaries within which they are allowed to move.

For several years, the EU has been a crucial participant in the field of digital technologies and trying to influence the behaviour of multinational companies. It sets regulations in cyber security, personal data protection and the use of AI, and expects global companies to adapt to the rules in place. To a certain extent, it is speculating and betting that the European market is still so large and important that leaving or not starting business cooperation would mean significant losses for investors. It must be added that this decoy game is currently working in Europe, but it is difficult to estimate how long the tactic will continue to work. It may happen that our own European potential starts to fade due to over-regulation. We must keep in mind that start-ups and entrepreneurs can always choose a region outside the EU to establish their company to avoid regulations, which would then deprive the EU of economic and technological power.

2.4 What Are the Gaps of the AI Act?

New technologies and innovations in AI will understandably bring challenges unforeseeable when the AI Act was originally formulated. For this reason, it will have to be regularly reviewed and updated in the future in order to be able to effectively regulate new systems. Yet, there are already significant gaps in the AI Act that can considerably complicate its upcoming implementation. Although the AI Act introduced the categorisation of individual systems, as mentioned above, they are formulated quite extensively. Thus, there are no clearly defined criteria to support these risk classifications. It is believed that most AI is mainly in the minimum-risk category, which means they will not need regulation. It is also hypothetical how sufficient oversight will be maintained to limit the development of AI within data and model control. The AI Act also does not include the impact of artificial intelligence on the environment and the need for public investment in AI.

3 A Balance for the Czech Republic

For the Czech Republic, finding a balance was extremely important. On one hand, there is a long tradition of human rights protection, closely associated with the consideration that the human being should always be at the centre of the approach to any new technology. On the other hand, the Czech Republic has a developing academic and business environment in AI, which would be a shame to jeopardise. Thus, finding a balance was crucial for the Czech Republic, and hopefully the fact that no one is entirely happy or unhappy is a sign that this balance has been achieved. It is also important to note that in the Czech Republic there are currently no specific laws directly regulating AI. It is expected that EU AI regulations will cover all member states, including Czechia. Despite this, Czechia is active in AI and will probably implement national regulations where permitted by EU law. The National AI Strategy from 2020, which is a component of the Czech Republic's 2019–2030 Innovation Strategy, supports the country's interest in AI advances and aims to establish the country as a leader in this area. Following on from the European Commission's proposal, the European Union's policy refers to using AI to boost global competitiveness.

As mentioned, the strategy aimed to use AI to establish the Czech Republic as a leader in this field among European nations. Research and development centred on AI and human-machine interaction in fields including manufacturing, mobility and security, as well as the establishment of a test facility and a European Centre of Excellence, are crucial components. Another key component is international cooperation, with the Czech Republic actively participating in global AI projects. The Ministry of Industry and Trade and the Deputy Prime Minister are responsible for assuring the accomplishment of the strategy's objectives.

In this regard, it is also necessary to discuss the Digital Decade 2030 initiative that establishes standards and oversight procedures for every member state's digital transformation by 2030. Through their national strategic plans, member states (including the Czech Republic) are required to report progress on a regular basis. The digital skills, digital infrastructure, digital transformation of enterprises, and digitalisation of public services are the four main areas of concentration in the Czech national plan. The objective is to assess the existing level of digital transformation and develop "goals to be achieved" trajectories based on KPIs. Areas of responsibility are divided between the relevant ministries and the recently formed Digital and Information Agency. Planning also includes collaboration on global initiatives, particularly via the EDIC consortium. With the first update scheduled for 2024 and subsequent updates every 2 years until 2030, the National Strategy will be modified on a regular basis in response to input from the European Commission.

3.1 And How Do the Czech MEPs See Things?

Czech MEPs do not agree on the rules for AI, and the Pirates party is the most annoyed. Even though 'casting' people using social scores, which they railed against, is prohibited, they are still concerned about the possible automated misuse of biometric data, such as camera footage. MEP Marcel Kolaja commented on the rules for AI on behalf of the Pirates. Although he agrees that rules for AI are needed, he stated that their form, which has emerged from negotiations with national governments, is a huge disappointment. "In fact, they pushed a directive into them that creates a legal framework for widespread surveillance of people with biometric cameras. With the help of artificial intelligence, they can recognise people's faces, tracking who they are when they were there, and with whom they were with. The rules should have banned this Orwellian tool; instead, they explicitly legalised it. That's an invasion of privacy that the Pirates really don't want to raise their hand for", Kolaja explained.

MEP and leader of SPOLU, candidate Alexandr Vondra (ODS) abstained from voting. "Although the final text negotiated with the member states is better than the EP's original position, it is still not optimal for the new standard to lead to the support of businesses dealing with artificial intelligence in the EU", Vondra stated. On the other hand, the people welcome the final wording of the act. According to MEP Michaela Šojdrová, the EU is thereby creating a similar precedent concurrently with the GDPR. According to her, it will be possible to use biometric identification systems in real-time, for example, for the targeted search of kidnapping victims or in the fight against terrorism. “It is this part that is key to finding missing children. Over 250,000 such children are registered in the EU every year. We must use the potential that these technologies offer and not just blindly ban its benefits”, Šojdrová said to explain her position. Her colleague and party member Tomáš Zdechovský has the same opinion: “AI tools can be a good servant but a bad master, so it is necessary to set development barriers that eliminate risks on the one hand. On the other hand, however, they will not hinder the further development of the prospective field”.

According to STAN MEP Stanislav Polčák, it is better to have rules that are insufficient in some parts than to have nothing at all. “The original ambition was diluted in the resulting agreement between the member states, the EP and the Commission. On the other hand, greater protection of human rights came into the final form”, he said. Concerned about scoring people, Luděk Niedermayer from TOP 09 also supported the Artificial Intelligence Act. “AI can speed up the use of today’s state of knowledge in practice, because it represents a very effective way of analysing available data and information”, the MEP noted. According to Niedermayer, the first months after AI’s “sharp launch” already displayed that AI will be used extensively in an attempt to influence attitudes in society, if not in political struggle. “From this point of view, it is a further evolution of techniques that have led some societies to a situation of great division and reduced ability to communicate between groups with different opinions. But this problem cannot be faced with bans and orders, instead with education, enlightenment, transparent politicians, and quality media”, Niedermayer declared.

The MEP of the ANO movement Dita Charanzová also agreed with the views for introducing rules on AI. She perceives excellent positives arsing from the use of AI, for example, in the field of cancer research and treatment. “During the deliberations, the aim for me was that the act should promote innovation in Europe and at the same time make the use of AI safe enough for citizens. For example, they will be sure that content created by artificial intelligence will be flagged”, Charanzová informed. However, Kolaja gave an example from practice where AI could be dangerous. “For example, I pushed through the legislation so that the rules apply to so-called e-proctoring, or programs for checking students for online exams. When this artificial intelligence is poorly trained, it can evaluate the noise from the hallway in the dormitory as cheating”, Kolaja explained. But according to him, there are more risks. “One of the biggest threats is the social score programs used by the Chinese government to caste people. We want to ban such use”, added Kolaja.

3.2 Obligations of Member States

It is necessary for each member state to supervise the observance of citizens' rights, oversee the launch of high-risk systems in state administration, and inform society about how these systems function. This step is relatively urgent and should be implemented in the next year. However, rushing the entire process is not advisable as setting it up to operate efficiently is complex. In the Czech Republic, there is currently a lively debate about which ministry or office should manage the implementation to maximise the use of existing structures and minimise costs.

4 National Strategies of Other EU Member States: Who Will Become a Leader in AI Innovations?

Member states have different approaches and priorities on AI in some areas, but in principle they have all been preparing for implementation of the recently introduced AI Act. Since 2018, the German AI federal strategy has significantly increased its investments in this region. In its annual report, the BMBF plans to invest as much as EUR 1.6 billion in AI. BMBF's current support includes findings, applications and a summary of 50 current initiatives aimed at improving productivity, building infrastructure and introducing practical technologies. The AI plan is a promising new impetus for the German AI ecosystem, the most comprehensive in the field, vision, production and maintenance. The key strategic plan is to transform consistent performance into concrete economic benefits, as well as providing clear guidance for companies. The proposed plan consists of several empirical studies on risks and advantages over the conventional risk management system, rather than classifying the risks associated with the system. The strategy also seeks to highlight that AI products labelled as "Made in Germany" or "Made in Europe" are currently too rare. The goal is to gradually establish an international trademark for high-quality and secure AI applications from Germany that serve the common good and align with fundamental European values.

Following the first AI for Humanity phase from 2018 to 2022, France's National Strategy for AI is a component of the France 2030 plan, which was enacted in November 2021. This plan is to speed up research and development for economic growth and increase France's pool of AI-trained professionals, which is a significant competitive advantage. Its main objectives are to raise national capabilities, establish France as a pioneer in reliable AI that is embedded, and further integrate AI into the economy through public and private partnerships. Developing French academic institutes into international hubs for AI and helping SMEs implement AI technology are also important foundations.

The Spanish National AI Strategy, published in December 2020, aims to support the development of AI across various sectors of the Spanish economy and society. Its main objectives include enhancing human capital in AI, fostering strong scientific excellence, and achieving a leading position in the development of tools and technologies for use of the Spanish language in AI. The strategy also includes measures to establish an ethical framework and infrastructure for AI, as well as initiatives addressing security and AI development to tackle societal challenges such as climate change and the COVID-19 pandemic. This national strategy is a flexible document that will be updated every 2 years, with coordination and monitoring overseen by the newly established entity SEDIA, supported by an advisory body on AI.

Similar to the Czech Republic, Germany and Spain, Italy is also trying to become a leader in AI. The Italian National AI Strategy from October 2020 aims to support the sustainable development of AI in Italy through several key measures. One of the main goals is to enhance AI education on all levels of the education system. This includes integrating AI courses into academic programmes and providing lifelong learning via online courses. Another important aspect of the strategy is to support AI research and innovation. Italy plans to increase investments in AI scientific research and establish new research laboratories, such as the PAI Lab in Pisa. These centres are intended to serve as bases for addressing scientific challenges and supporting AI innovations.

The national strategies of the above-mentioned countries are basically not that different, yet each assumes that it will reach as high as possible in this area. So, who will become the leader in AI within the European Union? The present state of AI development does not guarantee which country will become the innovation leader in this sector. Nevertheless, it is clear that important European leaders are working to ensure that the European Union as a collective takes a leading position in the field of AI.

5 What the New Legislation Will Protect Us From and What Will Be Completely Banned

The new legislation seeks to provide comprehensive protection against several documented dangers posed by artificial intelligence. These dangers include the potential misuse of AI for social control, mass surveillance, and discrimination. AI systems are often trained on vast amounts of both private and public data, which can reflect societal injustices and lead to biased outcomes, thereby exacerbating inequalities. From predictive policing tools to automated decision-making systems in the public sector, determining access to healthcare and social assistance, to monitoring the movements of migrants and refugees, AI can infringe on the human rights of the most marginalised in society. Other AI applications, like fraud detection algorithms, disproportionately affect ethnic minorities, leading to serious financial difficulties, as noted by Amnesty International.

5.1 Banned AI Practices

The legislation specifically targets surveillance systems and social scoring practices reminiscent of 'Big Brother'. For instance, social scoring systems, which assess individuals based on behaviour, driving habits, voting patterns, and more, are prohibited. Real-time tracking of individuals without clear justification is banned, with exceptions for serious crimes like murder, rape or terrorism. Yet, such systems cannot be used for general surveillance of ordinary citizens. Emotion recognition systems in workplaces and schools, used in some countries like China to monitor employee satisfaction or prevent cheating, are also banned due to their potential to undermine European values and personal freedoms.

The concept of 'Minority Report' AI systems is also relevant here. These AI systems assess an individual's risk of committing a crime based solely on profiling or personality traits. However, the legislation stipulates that such systems cannot be used to make standalone predictions about criminal behaviour. Instead, they are permitted only to assist human assessments that are grounded in objective and verifiable evidence directly related to criminal activities.

5.2 High-Risk AI Systems

High-risk AI systems include those used in significant life decisions, such as university admissions or employee productivity assessments, which can impact careers, education, and other life aspects. These systems require robust safeguards, including transparency about training data and decision-making processes, and human oversight to ensure fairness. Individuals must have the right to appeal decisions made by AI, access supervisory bodies, and contest outcomes. To avoid discrimination and bias, high-risk AI systems must use representative and complete datasets.

A notable example is the AI system used in the Netherlands to detect social benefit fraud, which failed due to poor data and lack of oversight, denying benefits to rightful claimants. Under the new legislation, such high-risk systems would require thorough documentation and human review to prevent similar errors and assure fair outcomes. The new AI legislation aims to balance the promotion of innovation with the protection of fundamental rights. It prohibits certain surveillance and social scoring systems while imposing stringent regulations on high-risk AI applications to ensure transparency, fairness and accountability. By setting these rules, the EU seeks to foster trust in AI technologies and prevent the misuse of AI that could bring about significant social harm.

5.3 Biometric Facial Recognition

Biometric identification systems are categorised in two groups: real-time biometric identification and ex post biometric identification from recordings. Real-time tracking is subject to much stricter rules and largely prohibited, with only three narrow exceptions. If you are walking down the street as an ordinary citizen and have not committed a crime or are not part of organised crime, no one will track you in real time. Cameras will not be used to monitor your movement. Still, there are exceptions for tracking terrorists or murderers, or locating missing or abducted children, where the use of these systems is crucial. These exceptions were advocated primarily by member states as they recognise the practical necessity in certain cases.

Biometric AI systems used for authentication or verification, such as those at airports for scanning travellers and allowing border passage, are generally permissible. The primary aim is to prevent the violation of personal freedom through real-time tracking, ensuring that no one knows your exact movements and actions without justified cause. At airports, these systems are activated when there is a suspicion of a terrorist or someone suspected of committing a serious crime. This use requires court intervention, preventing mass surveillance. The AI Act tackles the issue of biometric facial recognition with an emphasis on transparency and strict regulations to protect personal freedoms. While generative AI content must be clearly labelled, concerns about rights violations and illegal content dissemination are handled by other legal frameworks. Real-time biometric identification is heavily restricted, with allowances for specific, justified scenarios, ensuring protection against unwarranted mass surveillance.

5.4 Deep Fakes and Chatbots: Are They Still in the Limited Risk Category?

Deep fakes, altered videos and photos that are nearly impossible to distinguish from reality, are only partially addressed in the AI Act, chiefly from the perspective of generative AI. For non-high-risk systems, the Act mandates a certain level of transparency. Generative AI, which can create content that may not be entirely accurate, falls within this scope. For instance, when you see content on social media like X ('Twitter') or Facebook, the AI Act stipulates that if the content is generated by AI it should be labelled as such. Yet, the Act does not address whether the content infringes on rights or involves illegal dissemination. This falls under other legislation and national regulations concerning illegal content dissemination and violations of rights.

Chatbots are used to simulate human conversation in a wide range of applications, including education, information retrieval, commerce and services. They are easily available without the need for installation, which simplifies their use and integration with existing platforms. Providers of chatbots and digital assistants will now be required to ensure that their systems clearly inform end users that they are interacting with artificial intelligence. This information must be provided in an understandable form at the latest at the first contact with the technology. This will assure that users know they are interacting with an AI system and have a clear idea of how it can adapt to their interactions.

Deep fakes and chatbots are currently classified as "limited risk" AI systems in the AI Act. Yet, it should be remembered that their influence is already enormous in the world today, which raises the question of whether they are not underclassified. The legal framework for their performance is not clearly defined in the law, which focuses on the device for preventive measures rather than punitive ones; this includes, for example, technical guarantees against misuse.

5.5 Risk-Free Artificial Intelligence Systems

When it comes to risk-free systems, there is probably no need to discuss them extensively. Today, you can have smart calculators with some form of artificial intelligence, which the AI Act does not address. Then there are low-risk systems. A typical example is a chatbot used for purchasing tickets or handling booking and reservations. They are the simpler chats that do not have such a complex system that they can also discuss other than targeted topics. The rules for these low-risk systems mainly involve transparency. For instance, when you communicate through a chatbot, you must be informed that you are talking to a robot, not "Kate with blonde hair".

GPT chat models stand completely apart. This particular model was created by OpenAI, which is a conversational chatbot that you can ask anything and it will provide factual answers. However, it is a much more sophisticated chatbot than those used for purchasing plane tickets. And the biggest debates and controversies have centred on these systems where member states had to make the biggest compromises. So how did the final agreement on the regulation of Chat GPT systems turn out? The member states approved a common position at the end of 2022 during the Czech presidency, focusing more or less on the original proposal of the Commission, which did not include the regulation of multipurpose models like Chat GPT. By the end of 2022, discussions had shifted to the beginning of AI development, an area not addressed in the original proposal. The Commission did not deal with it, whereas the European Parliament only approved its position half a year later when Chat GPT was booming. The EP had the opportunity to focus more on the effects of GPT chats and their regulation. We can categorise models in two groups, but there must be clear transparency for all.

It is necessary to know how the models are trained. The models will also need technical documentation to detail the resources used and how the model works. There is a category of models with systemic risks that currently do not exist in Europe or globally. Yet, in a few years, it may include models like Google's Gemini or Chat GPT-5. These models are under development and, over time, will become very influential, potentially affecting society fundamentally. This means there is a potential risk, but it is unclear to what extent. Developers of these models should consider the risks their models may pose, think about possible impacts, how the model can be used, and try to prevent risks. The problem is that this is still abstract and has no practical impact because such models do not yet exist. We can imagine one possible impact using Chat GPT. When Chat GPT is used in education (writing a thesis, composing a poem, conducting research, writing a speech etc.), it is clear that it will have a huge impact due to the risk of fraud. On the other hand, anyone who has worked with Chat GPT knows that it is imperfect. Sometimes, the answers are incomplete or biased, perhaps

> There is a category of ChatGPT models with systemic risks that currently does not exist in Europe or globally.

against women or certain races. These systems have not been fully evaluated, meaning that the future impact is uncertain. It is not necessary to ban these systems outright, although it would be beneficial to have clear rules aligned with European values.

It is essential to know at least the source data from which the models learn. We should understand the models as semi-finished products, or some as products that will be sold to companies for further use. Therefore, technical documentation is appropriate and hence everyone who uses them knows how they work and what to do if a risk arises. A big question remains about the use of the system in law enforcement. This includes biometric identification for searching for criminals or using AI for predictive policing, which could evaluate whether someone is inclined to commit a crime. These are partly prohibited systems. AI could also be used in criminal analysis for browsing large data files, improving and speeding up criminal proceedings. This issue was important to the European Parliament, which wanted to ban a larger part of the system. In the end, a compromise was reached between law enforcement needs and model regulations. Since some models do not yet exist, there is a need for flexible regulation that can either become stricter or more relaxed to avoid hindering technological development. The risk with the regulation of these models is that they will work with much larger capacities, have more data, do many more things, and thus have a greater impact on how we learn, behave, create content, and work. They will be very capable models with great potential for change societal and business models. At the moment, it is difficult to predict potential problems that might arise from these models being launched and how to handle them.

Most of the rules based on the AI Act will come into force in approximately 2 years, around May 2026. There are a few exceptions and shortened deadlines, where greater urgency is indicated. The first is the case of prohibited practices, which will be enforced in half a year. For multipurpose AI models with a large impact, the rules for their regulation will start to apply 12 months after the AI Act comes into force, around August 2025.

6 Role of the European AI Office

Similar efforts will be made on the European level. Structures must be established to ensure a coordination mechanism across member states, preventing disparate actions and providing for consistent rules that facilitate compliance for businesses, developers, and system providers. Consultation with various interest groups, such as companies, non-governmental organisations, and trade unions, is crucial. Often, after legislative approval, implementation details are overlooked, yet these details are vital. Therefore, gathering extensive information makes sure that market needs and citizen protections are adequately addressed in both implementation and subsequent steps.

The European AI Office, to be established within the European Commission, will serve as the central hub for AI expertise across the EU, safeguarding the development and use of trustworthy AI while protecting against AI risks. It will implement the AI Act, particularly for general-purpose AI models, support governance bodies, enforce rules, and apply sanctions. The AI Office will promote an innovative ecosystem, supporting SMEs and start-ups through initiatives like 'GenAI4EU' and fostering international cooperation. Its structure will include five units and two advisors, focusing on excellence in AI. The European AI Office will oversee the AI Pact for business engagement and join the European AI Alliance for open policy dialogue.

6.1 Sanctions

As for high-risk systems and most of the obligations arising from the AI Act, fines for non-compliance can reach up to a maximum of EUR 15 million and 3% of the global turnover of a given company. For bans, the fines are even higher, reaching up to EUR 35 million or 7% of global turnover, with the higher amount always applying. It should be emphasised that each member state has room for adjustment to accommodate specific market situations. In other words, member states have the option to set a lower fine, for example, for a small and medium-sized enterprise or a start-up. They can also take into account whether the non-fulfilment of the obligation was deliberate, or simply an oversight.

7 AI and Jobs

The AI Act must assure that whenever an employer wants to deploy an AI system in the workplace, whether it in some production processes or some support systems, such as order management or supply chains, all of these systems should be communicated to employees in advance so they all know AI will be used in a given company in some way. AI will definitely change the job market, the nature of our work and the composition of the labour market. It will also affect what will be required of employees. What can a member state do about it? The same as always when a new technology emerges: inform citizens and companies about it, and help them prepare by retaining workers, for example. However, it is important to note that, based on the OECD report from 2023, 27% are at risk due to automation in the European Union, while for example the Czech Republic is even above the average with 35%.

There are also voices calling for new legislation to address the impact of AI on the labour market in order to protect employees. Whether this actually happens, how strict the rules might be, and what they will look like remains uncertain. Interestingly, AI and its latest developments hold the potential to replace white-collar jobs rather than those involving manual labour. It is said that the profession least threatened by AI is the florist. While AI might be able to write a better speech, article, book or diploma today, we still lack the technology for AI to replicate the human touch necessary to replace or even threaten jobs such as florists or those involving sensitive manual labour.

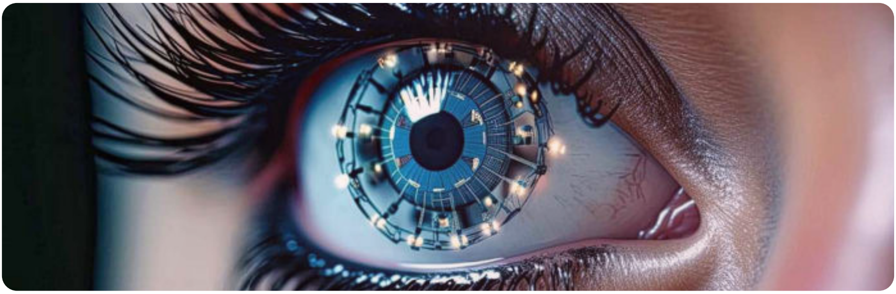
8 Conclusion

Last year was marked by enthusiasm with AI, but at the same time it led legislators and the public to discuss the safety and regulation of this technology. A hectic year in technology started with the launch of ChatGPT at the end of 2022 and ended with the landmark agreement on the EU AI Act. Early signs suggest that this first 'rulebook for artificial intelligence' in the Western world goes some way to protecting people from harm caused by AI, but it is still insufficient in a number of crucial areas and does not ensure the protection of human rights, especially for the most marginalised. The key question for 2024 is hence whether all discussions on the regulation of AI will lead to concrete commitments, bring solutions to current risks and find a place in laws. But what makes regulating AI complex and challenging? The first problem is the vague nature of the term "artificial intelligence" itself, which makes efforts to regulate the technology challenging. There is no universal agreement on a definition as the term does not refer to one specific technology and instead encompasses a myriad of technological applications and methods. In addition, the use of AI systems in multiple different areas across the public and private sectors means that a large number of diverse stakeholders are involved in their development and implementation, and thus these systems are a complex product of labour, data, software and financial inputs, and any regulation must contend with antecedent and consequential damages.

Parallel to the EU legislative process, the UK, USA and other countries have set out different plans and approaches to identify the key risks posed by AI technologies and how they intend to mitigate them. Although there are many obstacles in these legislative processes, they should not limit any efforts to protect people from current and future harms caused by AI. Regulation must be legally binding and focus on already documented human-caused harm. Likewise, any regulation must include broader accountability mechanisms beyond the technical assessments that industry enforces. While these can be a useful part of any regulatory policy, especially when testing algorithmic bias, necessary prohibitions and restrictions cannot be ruled out, especially for systems that are questionable in terms of human rights, no matter how accurate or technically efficient they seem to be.

Others must learn from the EU's approach and assure there are no legal loopholes that would allow private or public entities to circumvent regulatory obligations in any way. It is also important that in cases where future regulation restricts or bans the use of certain AI systems within one piece of legislation, there are no legal avenues to allow the same systems to be exported to other countries where they could be used to harm human rights.

With the arrival of 2024, the time has come not only to ensure that respect for the law is already in mind when AI systems are created, but also to make sure that those affected by this technology are involved in deciding how it should be regulated and that their experiences were taken into account in the discussions.



REFERENCES

- Adamopoulou E. & Moussiades L. (2020). An overview of Chatbot Technology. Artificial Intelligence Applications and Innovations. 2, 373-383. Retrieved 10 July 2024 from: https://link.springer.com/chapter/10.1007/978-3-030-49186-4_31?ref=blog.min.io
- Blumenová, K. (2024). Zákaz sociálního skóringu i hromadného rozpoznávání obličeju. Co přinese Evropská regulace umělé inteligence? Retrieved 11 July 2024 from: https://www.hrot24.cz/clanek/zakaz-socialniho-skoringu-i-hromadneho-rozpoznavani-obliceju-evropsky-ai-act-je-prvni-skutecnou-regulaci-ai
- BMBF. (2023). Aktionplan ,,Kunstliche Intelligenz". Retrieved 11 July 2024: https://www.bmbf.de/bmbf/de/forschung/digitale-wirtschaft-und-gesellschaft/kuenstliche-intelligenz/ki-aktionsplan.html
- CNN. (2023). Top companies raise alarm over Europe´s proposed AI law. Retrieved 10 July 2024 from: https://edition.cnn.com/2023/06/30/tech/eu-companies-risks-ai-law-intl-hnk/index.html
- Edwards, L. (2022). Expert Opinion: Regulating AI in Europe. Retrieved 10 July 2024 from: https://www.adalovelaceinstitute.org/report/regulating-ai-in-europe/
- EURACTIV. (2024). The AI Act vs. deepfakes: A step forward, but is it enough?. Retrieved 10 July 2024 from: https://www.euractiv.com/section/artificial-intelligence/opinion/the-ai-act-vs-deepfakes-a-step-forward-but-is-it-enough/
- European Civic Forum. (2024). Packed with loopholes: Why the AI Act fails to protect civic space and the rule of law. Retrieved 10 July 2024 from: https://civic-forum.eu/advocacy/artificial-intelligence/packed-with-loopholes-why-the-ai-act-fails-to-protect-civic-space-and-the-rule-of-law
- European Commission a. (2024). AI Act. Retrieved 10 July 2024 from: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai KPMG. (2023). EU Artificial Intelligence Act. Retrieved 10 July 2024 from: https://assets.kpmg.com/content/dam/kpmg/ie/pdf/2024/01/ie-eu-artificial-intelligence-act.pdf
- European Commission b. (2024). European AI Office. Retrieved 10 July 2024 from: https://digital-strategy.ec.europa.eu/en/policies/ai-office
- European Commission c. (2023) France - National Strategy for AI. Retrieved 11 July 2024: https://digital-skills-jobs.europa.eu/en/actions/national-initiatives/national-strategies/france-national-strategy-ai
- Hacker, P. (2023). What's Missing from the EU AI Act. Retrieved 10 July 2024 from: https://verfassungsblog.de/whats-missing-from-the-eu-ai-act/
- Holzman, O. (2024). Regulace umělé inteligence má zamezit zneužití, ale nesmí zastavit rozvoj, říka přední muž Microsoftu. Retrieved 11 July 2024: https://cc.cz/regulace-umele-inteligence-ma-zamezit-zneuziti-ale-nesmi-zastavit-rozvoj-rika-predni-muz-microsoftu/
- Kreč, L. a. (2024). Žádná umělá inteligence od Applu, žádné upravené plodiny. Opravdu evropská regulace zlepšuje životy?. Retrieved 10 July 2024 from: https://cc.cz/zadna-umela-inteligence-od-applu-zadne-upravene-plodiny-opravdu-evropska-regulace-zlepsuje-zivoty/

- Kreč, L. b. (2024). Úchvatná cesta českého vědce na vrchol: Měl zkoumat pomeranče, otrávil se a pak objasnil růst květin. Retrieved 10 July 2024 from:
- Ministerstvo obchodu a prumyslu (2019.) Národní strategie umělé inteligence v České republice. Retrieved 11 July 2024: https://vlada.gov.cz/assets/evropske-zalezitosti/umela-inteligence/NAIS_kveten_2019.pdf
- Nejedlý, M. (2024). Europoslanci schválili pravidla pro AI. Piráti se bojí čínskych praktik. Retrieved 10 July 2024 from: https://www.idnes.cz/zpravy/domaci/evropsky-parlament-umela-inteligence-regulace-hlasovani-ai-act.A240313_115802_domaci_nema
- OECD. (2023). OECD Employment Outlook 2023. Retrieved 10 July 2024 from: https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en.htm
- Stănilă, M. L. (2020). Minority report: AI Criminal Investigation. Towards a Better Future: Human Rights, Organized Crime and Digital Society, 1, 142-155. Retrieved 10 July 2024 from: https://eprints.uklo.edu.mk/id/eprint/5884/1/KICEVO_2020.pdf#page=142
- The Economist Newspaper. (2023). Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. The Economist. Retrieved 10 July 2024 from: https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation
- Trachtová, Z. Evropa připravila regulaci umělé inteligence. ‚Velký bratr v EU nebude,' tvrdí expertky. (2024). iROZHLAS. https://www.irozhlas.cz/zpravy-svet/eu-umela-inteligence-regulace-evropa-brusel_2402120727_ako
- White&Case. (2024) AI Watch: Global Regulatory Tracker - Czech Republic. Retrieved 10 July 2024 from: https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-czech-republic

Chapter 9

The Regulation of AI in Research by Adapting EU AI Regulations to African Contexts and Exploring Opportunities and Challenges for Harmonious AI Governance

David Mhlanga

1 Introduction

Background: The Importance of AI in Modern Research

Artificial Intelligence (AI) has become a central pillar of contemporary research, revolutionising various fields by enhancing data analysis capabilities, enabling advanced simulations, and fostering innovative methodologies (Salih& Mhlanga 2023; Mhlanga 2023). AI's ability to process vast amounts of data at unprecedented speeds allows researchers to uncover patterns, make predictions, and generate previously unattainable insights. For instance, AI algorithms are used in healthcare to predict disease outbreaks, personalise treatments, and improve diagnostic accuracy (Topol, 2019).

Similarly, in environmental science, AI aids in climate modelling and the analysis of satellite imagery to monitor environmental changes (Rolnick et al., 2019). The economic and social sciences have also benefited from AI, with machine learning models providing deeper insights into consumer behaviour, market trends, and social dynamics (Agrawal, Gans, & Goldfarb, 2018). Despite these advancements, the rapid integration of AI into research poses considerable ethical, legal and social challenges. These include concerns about data privacy, algorithmic bias, and the accountability of AI systems. As AI continues to evolve, the need for robust and comprehensive regulatory frameworks becomes ever more critical to ensure that its research deployment is ethical, transparent and beneficial to society (Mhlanga & Salih 2023, Mhlanga 2024).

Globally, different regions have adopted varied approaches to AI regulation, reflecting their unique legal, cultural and economic contexts. The European Union (EU) has been a leader in developing comprehensive AI regulations to provide for ethical AI deployment, protect fundamental rights, and promote trustworthy AI. The EU's proposed Artificial Intelligence Act, which occurred on the first of August 2024, classifies AI systems based on risk levels. It imposes stringent requirements on high-risk applications, including those used in critical infrastructure, education, and employment (European Commission, 2021). In contrast, the United States has taken a more decentralised and sector-specific approach to AI regulation, focusing on innovation and economic competitiveness. This approach involves multiple agencies issuing guidelines tailored to sectors such as healthcare, transportation and finance (Executive Office of the President, 2020). China, meanwhile, has implemented a combination of regulatory measures that stress both the strategic importance of AI for economic growth and the need for state control and oversight (Ding, Triolo, & Sacks, 2018). Understanding these diverse regulatory landscapes is essential for developing effective AI governance frameworks that can be adapted to different regional contexts, particularly in Africa. African nations face unique challenges and opportunities regarding AI adoption, calling for a tailored approach to regulation that considers local socio-economic and technological conditions.

This chapter examines the critical issue of regulating AI within the research sphere, concentrating on adapting the European Union's AI regulations to African contexts. The purpose is to explore how these regulations can be modified to address African nations' special challenges and opportunities. By analysing the EU's approach and identifying the adaptations needed, the chapter proposes a framework for inclusive and adaptable AI governance that ensures ethical use and promotes innovation. In addition, the chapter aims to highlight the potential barriers to implementation and suggest strategies to overcome these challenges. It emphasises the importance of international collaboration and local stakeholder engagement in creating effective and contextually relevant regulations. Through comprehensive analysis, the chapter aspires to contribute to the ongoing dialogue on effective AI governance that respects cultural diversity while fostering global technological harmony. The central problem considered in this chapter is the challenge of adapting the European Union's AI regulations to African countries' diverse and complex socio-economic, cultural and technological landscapes. While the EU has taken significant strides in establishing a comprehensive AI regulatory framework, these regulations are designed with the EU's specific context in mind. Transposing these regulations to African contexts without modifications may lead to ineffective governance, stifled innovation, and unaddressed ethical concerns. This means there is a critical need to develop a customised approach that harmonises AI governance across regions, assuring that AI's benefits are equitably distributed while mitigating potential risks.

2 The EU's Approach to AI Regulation

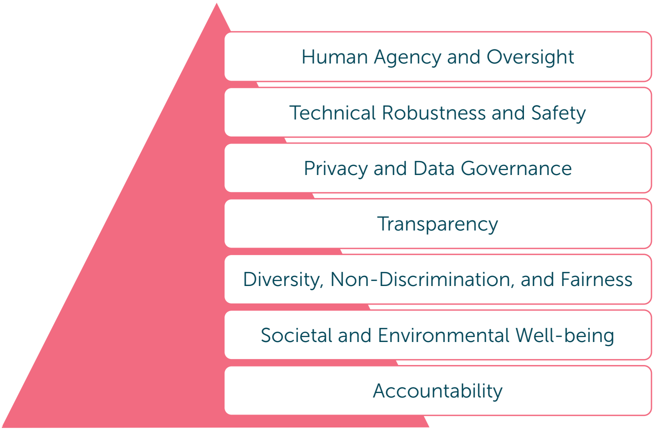
2.1 Key Directives and Guidelines Governing AI in the EU

The European Union has proactively developed a comprehensive regulatory framework for AI. The cornerstone of this effort is the European Commission's proposal for the Artificial Intelligence Act(in force 1.8.2024), which aims to create a legal framework that addresses the risks and challenges specific to AI while fostering innovation and investment in AI technology. This act categorises AI applications on three risk levels: unacceptable risk, high risk, and low or minimal risk (European Commission, 2021). Unacceptable risk applications, such as those violating fundamental rights, are prohibited. High-risk applications, which include AI systems used in critical infrastructures, education, employment, and law enforcement, are subject to stringent requirements, including conformity assessments and post-market monitoring. Low or minimal risk applications encompass most AI systems and are encouraged to adhere to voluntary codes of conduct and ethical guidelines. Moreover, the EU established the General Data Protection Regulation (GDPR) as a fundamental framework that indirectly impacts AI by emphasising data protection, privacy, and individual rights. The EU's Ethics Guidelines for Trustworthy AI, developed by the High-Level Expert Group on AI, further outline principles for making sure AI systems are lawful, ethical and robust (European Commission, 2018; High-Level Expert Group on Artificial Intelligence, 2019).

Objectives and Principles Behind the EU AI Regulations

Several core objectives and principles drive the EU's AI regulations. The primary objective is to ensure that AI systems are safe and respect existing laws, including fundamental rights. This encompasses protecting individuals from potential harm and biases that AI systems might introduce. The principles behind the EU's AI regulations are listed in Figure 1 below.

Figure 1: Objectives and Principles Behind the EU AI Regulations


Human Agency and Oversight
Technical Robustness and Safety
Privacy and Data Governance
Transparency
Diversity, Non-Discrimination, and Fairness
Societal and Environmental Well-being
Accountability

Source: Author (2024)

The principles underlying the EU's AI regulations stress several key areas to provide for ethical and responsible use of AI technology. First, the principle of human agency and oversight ensures that AI systems are designed to augment human capabilities and decisions, maintaining meaningful human control and oversight over these systems. This principle underscores the importance of humans being able to intervene and oversee AI operations to assure they align with human values and intentions. Second, technical robustness and safety are paramount. AI systems must be reliable, secure and resilient to attacks or manipulations, providing for their safe operation throughout their lifecycle. This includes rigorous testing and validation to prevent malfunctions and vulnerabilities that could be exploited. Privacy and data governance form another critical pillar. AI systems must comply with privacy and data protection regulations, such as the GDPR, ensuring the protection of personal data and the privacy of individuals. These principal mandates require strict adherence to data protection laws, safeguarding user information from misuse and unauthorised access. Transparency is another key principle. The operations of AI systems should be explainable, traceable and transparent, allowing users to understand how these systems function and make decisions. This transparency fosters trust and enables users to hold AI systems accountable for their actions and decisions.

Further, diversity, non-discrimination and fairness are crucial in AI design. AI systems should be developed to avoid biases and ensure inclusivity, promoting equal opportunities and preventing discrimination. This principle aims to mitigate the risk of AI perpetuating or amplifying existing societal biases and inequalities. The societal and environmental well-being principle underscores the positive impact AI should have on society. AI technologies should contribute to enhancing social well-being and promoting sustainability. This includes developing AI applications that address societal challenges and environmental issues, contributing to a better and more sustainable future. Finally, accountability is essential. Clear mechanisms must be in place to assure accountability for AI systems and their outcomes. This involves establishing clear guidelines and responsibilities for AI system developers, operators and users, ensuring that any adverse effects or malfunctions are addressed and rectified promptly. These principles collectively aim to create a framework for the ethical and responsible development and deployment of AI technologies, to make sure they benefit society while minimising potential risks and harms (European Commission, 2018).

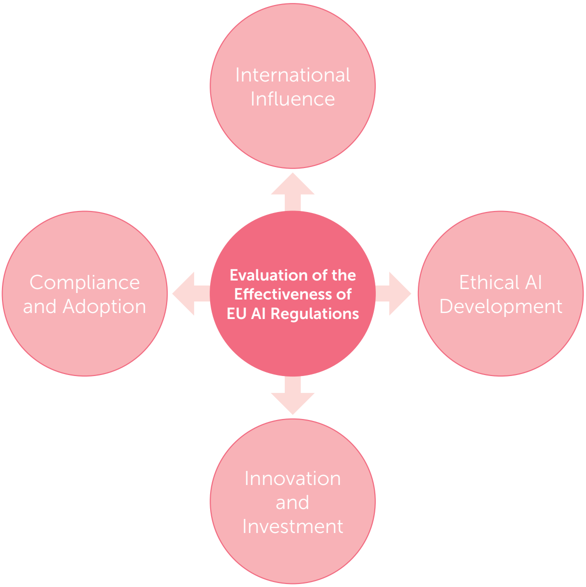
3 AI Regulation Implementation Within EU Member States

Several EU member states have begun implementing the principles and guidelines established by the EU on the national level. For instance, Germany adopted the AI Strategy focusing on making AI a vital driver of economic growth and innovation while ensuring ethical standards. The strategy emphasises research, investment in AI infrastructure, and international collaboration (Germany Federal Ministry for Economic Affairs and Energy, 2018). France also launched its AI for Humanity strategy, which aligns with the EU's principles by prioritising ethical AI development, transparency, and public trust. France's approach includes investing in AI research, fostering public-private partnerships, and making sure AI systems are designed with inclusivity and fairness (Government of France, 2018). Another notable example is Finland, which implemented the National AI Strategy to integrate AI into various sectors, including healthcare, education, and public administration. Finland's strategy focuses on AI literacy, ethical guidelines, and regulatory frameworks to ensure responsible AI deployment (Ministry of Economic Affairs and Employment of Finland, 2017).

4 Evaluation of the Effectiveness of EU AI Regulations

The effectiveness of the EU's AI regulations can be evaluated through the lenses listed in Figure 2 below.

Figure 2: Evaluation of the Effectiveness of the EU AI Regulations


International Influence
Compliance and Adoption
Evaluation of the Effectiveness of EU AI Regulations
Ethical AI Development
Innovation and Investment

Source: Author (2024)

An evaluation of the effectiveness of the EU's AI regulations can be made via several lenses. First, compliance and adoption indicate a positive trend among AI developers and organisations within the EU. Initial assessments show that the mandatory requirements for high-risk AI applications have driven significant investments to make sure AI systems meet regulatory standards. Second, the impact on innovation and investment is notable. The EU's balanced approach to regulation, combining mandatory requirements with innovation support, has created an environment conducive to AI development. There has been substantial growth in AI research funding and public-private partnerships, suggesting the regulations have not stifled innovation. Ethical AI development is another critical aspect. The emphasis on ethical guidelines and principles has promoted the development of AI systems that prioritise fairness, transparency and accountability. This focus has helped build public trust in AI technologies, which is essential for widespread adoption. Moreover, the EU's regulatory framework has had a significant international influence. It has set a global standard for AI governance, prompting other regions to adopt similar principles. The EU's leadership in AI regulation positions it as a key player in the global discourse on ethical AI. However, challenges remain in ensuring uniform implementation across all member states and addressing the rapid pace of AI advancements. Continuous monitoring, stakeholder engagement, and iterative improvements to the regulatory framework are required to maintain its effectiveness (European Commission, 2018).

5 Adapting EU AI Regulations to African Contexts

5.1 Comparative Analysis of Socio-Economic, Cultural and Technological Landscapes Between the EU and African Nations

The European Union (EU) and African nations vary considerably in their socio-economic, cultural and technological landscapes. High levels of economic development, advanced technological infrastructure, and well-established legal frameworks characterise the EU. In contrast, many African nations face challenges such as lower levels of economic development, less developed technological infrastructure, and varied legal systems. Economically, the EU benefits from a highly integrated market and substantial investment in technology and innovation. In African countries, economic conditions vary widely, with some nations experiencing rapid growth and others facing persistent poverty and underdevelopment. This economic disparity impacts the ability of African nations to invest in and regulate emerging technologies like AI (European Commission, 2021). Culturally, the EU tends to have more homogeneous regulatory environments, whereas African nations exhibit a range of cultural and regulatory practices. This diversity calls for a more flexible approach to AI regulation to accommodate cultural norms and values (High-Level Expert Group on Artificial Intelligence, 2019). Technologically, the EU has widespread access to advanced digital infrastructure, high Internet penetration rates, and robust research and development capabilities. In contrast, many African countries struggle with limited digital infrastructure, lower Internet penetration, and fewer resources dedicated to technological research. These technological gaps pose significant challenges to adopting and regulating AI technologies in Africa (Germany Federal Ministry for Economic Affairs and Energy, 2018; Ministry of Economic Affairs and Employment of Finland, 2017).

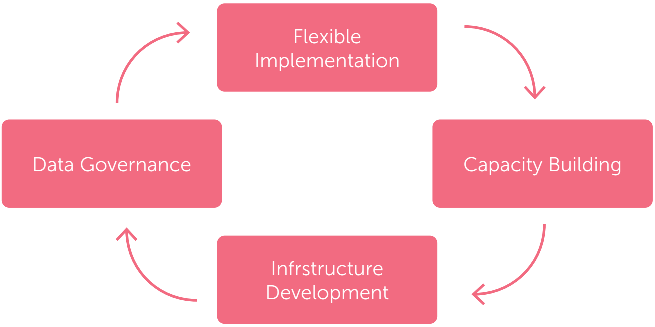
5.2 Identification of Unique Challenges in the African Context

African nations encounter several unique challenges in adopting and regulating AI technologies. A primary challenge is the lack of infrastructure. Many African regions have limited access to reliable electricity, Internet connectivity, and digital devices, which hinders the deployment and utilisation of AI technologies (Government of Kenya, 2021). Another big challenge is the shortage of skilled professionals. The development and regulation of AI require expertise in various fields, including computer science, data science, and ethics. However, many African countries face a shortage of professionals with the necessary skills to develop, implement and regulate AI systems (South African Government, 2020). Data privacy and security are also critical concerns. Many African countries lack comprehensive data protection laws, raising fears regarding the ethical use of AI and protecting personal information. The lack of standardised data governance frameworks further complicates efforts to provide for the ethical deployment of AI technologies (Nigerian Government, 2021). In addition, socio-economic disparities impact the equitable distribution of AI benefits. AI can potentially exacerbate existing inequalities if not carefully managed, particularly in regions with high levels of poverty and unemployment (European Commission, 2018).

6 Potential Modifications to EU Regulations to Better Fit African Contexts

Several modifications are needed to adapt EU AI regulations to the African context, as shown in Figure 3 below.

Figure 3: Potential Modifications to EU Regulations to Better suit African Contexts



Source: Author (2024)

First, regulations should be flexible to accommodate African countries' diverse economic, cultural and technological conditions. This could involve creating tiered regulatory requirements based on a country's level of technological development and AI-readiness (High-Level Expert Group on Artificial Intelligence, 2019). Second, investing in education and training programmes to develop a skilled workforce capable of managing AI technologies is fundamental. Partnerships with international organisations and universities can help build local expertise in AI (South African Government, 2020). Infrastructure development is another essential aspect. Supporting the development of digital infrastructure is critical for the effective deployment of AI. International aid and investment can significantly contribute to building the necessary infrastructure (Government of Kenya, 2021). Contextual ethical guidelines should also be considered. Ethical guidelines must be adapted to reflect local cultural values and societal norms. This could involve engaging local communities in developing ethical standards for AI (European Commission, 2018). Establishing robust data protection laws and governance frameworks tailored to the African context is also vital. This includes creating policies that protect personal data and ensure transparency in AI decision-making processes (Nigerian Government, 2021). Lastly, encouraging regional collaboration among African countries can help harmonise AI regulations and promote the sharing of best practices. Regional bodies such as the African Union can play a key role in coordinating these efforts (Government of Rwanda, 2019).

6.1 Examples of Existing AI Regulatory Frameworks in African Countries

While comprehensive AI regulatory frameworks are still emerging in Africa, several countries have already made strides in this area. South Africa has developed a National Artificial Intelligence Strategy that outlines the country's approach to AI, focusing on ethical guidelines, capacity-building, and research and development. The strategy emphasises the importance of AI in driving economic growth and addressing societal challenges (South African Government, 2020). Kenya has proactively adopted digital technologies and initiated efforts to develop AI policies. The government has recognised the potential of AI in various sectors, including agriculture, healthcare and education, and is working towards creating an environment conducive to AI innovation (Government of Kenya, 2021). Nigeria has also shown interest in AI development and regulation. The country is concentrating on building digital infrastructure and creating policies supporting innovation while ensuring

AI technologies' ethical use (Nigerian Government, 2021). Rwanda has positioned itself as a hub for technological innovation in Africa. The country has implemented policies to promote digital transformation and is exploring the development of AI regulations that align with its vision of becoming a knowledge-based economy (Government of Rwanda, 2019). These examples illustrate the growing interest in AI regulation across the continent and the efforts being made to create frameworks that address the unique challenges and opportunities in African countries.

6.2 Adapting EU AI Regulations to African Contexts

Comparative Analysis of Socio-Economic, Cultural and Technological Landscapes Between the EU and African Nations

The EU and African nations differ significantly in socio-economic, cultural, and technological landscapes. High economic development, advanced technological infrastructure, and well-established legal frameworks characterize the EU. At the same time, many African nations face challenges like lower economic development, less developed technology, and varied legal systems (European Commission, 2021). The EU benefits from an integrated market and strong technology investment, whereas African countries face economic disparity, limiting AI investment (European Commission, 2018). Culturally, Africa's diversity necessitates flexible AI regulation, unlike the EU's more homogeneous practices (High-Level Expert Group on Artificial Intelligence, 2019). Technologically, the EU has better digital infrastructure and research capabilities, while Africa struggles with limited resources and Internet access (Germany Federal Ministry for Economic Affairs and Energy, 2018; Ministry of Economic Affairs and Employment of Finland, 2017).

Unique Challenges in the African Context

African nations face key challenges in adopting and regulating AI technologies. Limited infrastructure, including unreliable electricity and Internet access, restricts AI deployment (Government of Kenya, 2021). There is also a shortage of skilled professionals to develop and regulate AI (South African Government, 2020). Additionally, concerns about data privacy and security are heightened due to the lack of comprehensive data protection laws and standardized governance frameworks (Nigerian Government, 2021). Socio-economic disparities further complicate AI adoption, as unequal access to AI could worsen poverty and unemployment (European Commission, 2018).

Examples of Existing AI Regulatory Frameworks in African Countries

While comprehensive AI regulatory frameworks are still developing in Africa, several countries have made notable progress. South Africa's National Artificial Intelligence Strategy emphasizes ethical guidelines, capacity-building, and AI-driven economic growth (South African Government, 2020). Kenya has adopted digital technologies and is working on AI policies, recognizing AI's potential in agriculture, healthcare, and education (Government of Kenya, 2021). Nigeria focuses on building digital infrastructure and supporting AI innovation while ensuring ethical use (Nigerian Government, 2021). Rwanda aims to become a technological hub by promoting digital transformation and developing AI regulations (Government of Rwanda, 2019). These efforts reflect the growing interest in AI regulation across the continent.

7 Opportunities for Harmonious AI Governance

7.1 Synergies Between EU and African Regulatory Approaches

The regulatory approaches of the EU and African nations, while distinct in their contexts, offer significant synergies that could be harnessed to create a more effective and harmonious AI governance framework. The EU's emphasis on ethical AI development, transparency and accountability provides a robust foundation for African countries to adapt to their specific needs. African nations, on the other hand, bring diverse cultural perspectives and unique societal needs that can enrich the ethical considerations and inclusivity of AI governance. By integrating these approaches, it is possible to develop a regulatory framework that is both comprehensive and adaptable, leveraging the strengths of both regions (European Commission, 2021; High-Level Expert Group on Artificial Intelligence, 2019).

7.2 Benefits of Harmonising AI Governance Across Diverse Regions

Harmonising AI governance across diverse regions, such as the EU and Africa, offers several benefits. First, it promotes consistency in AI regulations, facilitating cross-border collaboration and innovation. Consistent regulatory standards help businesses and researchers navigate legal requirements more easily, reducing barriers to entry and encouraging a more collaborative international AI ecosystem. Second, harmonised regulations ensure that AI technologies developed and deployed in different regions adhere to common ethical standards. This is crucial for addressing global challenges such as data privacy, algorithmic bias, and the ethical use of AI. By adopting a unified approach, regions can collectively work towards mitigating the risks associated with AI and making sure that its benefits are distributed equitably. Finally, harmonised AI governance can enhance trust among stakeholders, including governments, businesses, and the public. When transparent and consistent regulations govern AI systems, confidence is built in their safety, reliability, and ethical integrity. This trust is essential for the widespread adoption and acceptance of AI technologies (European Commission, 2018).

7.3 Opportunities for Fostering Innovation Through Tailored AI Regulations

Customised AI regulations considering different regions' unique socio-economic and cultural contexts can foster innovation in several ways. In Africa, for example, regulations prioritising local challenges and opportunities can spur the development of AI solutions targeting specific needs, such as improving healthcare access, enhancing agricultural productivity, and addressing educational disparities. Flexible regulatory frameworks that support experimentation and pilot projects can encourage innovation by allowing businesses and researchers to explore new AI applications without the burden of stringent compliance requirements. This approach can help identify best practices and inform the development of more effective regulations over time. In addition, tailored regulations can promote inclusive innovation by ensuring that marginalised communities can access AI technologies and their benefits. By incorporating principles of fairness and equity into AI governance, regions can foster innovations that address social inequalities and contribute to sustainable development (South African Government, 2020; Government of Kenya, 2021).

7.4 Potential for International Collaborations and Partnerships

International collaborations and partnerships are crucial for advancing AI governance and leveraging the benefits of AI technologies globally. The EU and African nations can collaborate on various fronts, including research and development, policymaking, and capacity-building. Collaborative research initiatives can pool resources and expertise from different regions to address common challenges and develop innovative AI solutions. For instance, partnerships between European and African universities and research institutions can facilitate knowledge exchange and drive advancements in AI technology. Policy dialogues and exchanges can help harmonise AI regulations and share best practices. By engaging in international forums and working groups, regions can collaboratively align their regulatory approaches and address global AI governance issues. Capacity-building initiatives can support the development of skilled professionals in AI across different regions. Joint training programmes, workshops, and scholarships can boost AI literacy and expertise, especially in regions with limited resources. These collaborations can also help build the institutional capacity to implement and enforce AI regulations effectively (Nigerian Government, 2021; Government of Rwanda, 2019). In summary, the synergies between EU and African regulatory approaches, the benefits of harmonising AI governance, the opportunities for fostering innovation via customised regulations, and the potential for international collaborations and partnerships highlight the substantial opportunities for creating a harmonious and effective global AI governance framework.

7.5 Challenges and Barriers to Implementation

Resource Limitations in African Nations

Resource limitations are a significant barrier to the effective implementation of AI regulations in many African countries. These limitations include financial constraints, limited access to advanced technology, and insufficient human resources. Financial constraints often hinder investment in the infrastructure required for AI development and deployment. Many African nations struggle with budget limitations that affect their ability to invest in cutting-edge technology and comprehensive regulatory frameworks. Moreover, there is often a lack of access to advanced technology and research facilities, which impedes the growth and integration of AI within these countries (South African Government, 2020).

Infrastructural Disparities and Their Impact on AI Deployment

Infrastructural disparities considerably impact the deployment and effectiveness of AI technologies in Africa. Many regions lack reliable electricity, Internet connectivity, and the digital infrastructure to support AI systems. In countries where these basic infrastructural elements are not consistently available, deploying AI technologies becomes a substantial challenge. This disparity hinders the development and utilisation of AI and affects the ability to implement and enforce AI regulations effectively. The digital divide between urban and rural areas exacerbates these challenges, with rural regions often being the most underserved (Government of Kenya, 2021).

Varying Levels of AI Readiness Across African Countries

The levels of AI readiness vary significantly across African countries, influenced by economic development, education systems, and technological infrastructure. Countries like South Africa and Kenya have made notable progress in AI readiness, developing strategies and frameworks to support AI innovation and deployment. However, many other African nations are still in the nascent stages of AI adoption, lacking the necessary infrastructure, expertise, and regulatory frameworks. This variability makes implementing a standardised approach to AI regulation across the continent challenging. It calls for tailored strategies that consider each country's specific context and readiness (Government of Rwanda, 2019).

Addressing Ethical Concerns and Ensuring Equitable Benefits

Addressing ethical concerns and ensuring equitable benefits from AI technologies are critical challenges in Africa. Issues like data privacy, algorithmic bias, and the potential for AI to exacerbate existing social inequalities must be carefully managed. Many African countries lack comprehensive data protection laws, raising concerns about the ethical use of AI and protecting personal information. Ensuring that AI technologies are designed and deployed in ways that promote fairness and inclusivity is essential to avoid reinforcing existing disparities. Further, it is vital to ensure that the benefits of AI are equitably distributed, providing opportunities for all segments of society to benefit from technological advancements (Nigerian Government, 2021).

Strategies to Overcome These Challenges

Several strategies can be employed to overcome these challenges and facilitate the effective implementation of AI regulations in Africa. Capacity-building investing in education and training programmes to develop a skilled workforce capable of managing AI technologies is crucial. Partnerships with international organisations and universities can assist in building local expertise in AI, assuring enough qualified professionals to develop, implement and regulate AI systems. Infrastructure Development: supporting the development of digital infrastructure is essential for the effective deployment of AI. International aid and investment can significantly help build the infrastructure needed, such as reliable electricity and Internet connectivity, particularly in underserved regions. Tailored Regulatory Frameworks: developing flexible and context-specific regulatory frameworks that consider the unique challenges and opportunities in different African countries can help ensure that AI regulations are effective and relevant. This includes creating tiered regulatory requirements based on a country's technological development and AI readiness.

Ethical Guidelines and Standards: establishing robust ethical guidelines and standards tailored to the African context is vital. This involves engaging local communities in developing ethical standards for AI to ensure that they reflect local cultural values and societal norms. Comprehensive data protection laws and governance frameworks are also necessary to protect personal data and ensure transparency in AI decision-making processes. Regional Collaboration: Encouraging regional collaboration among African countries can help harmonise AI regulations and promote the sharing of best practices. Regional bodies such as the African Union can be key in coordinating these efforts, facilitating policy dialogues, and supporting capacity-building initiatives. By addressing these challenges through targeted strategies, African nations can overcome the barriers to AI implementation and create an environment conducive to AI innovation and regulation.

> ”
> Comprehensive data protection laws and governance frameworks are necessary to protect personal data and ensure transparency in AI decision-making processes.

8 Proposed Framework for Inclusive and Adaptable AI Regulation

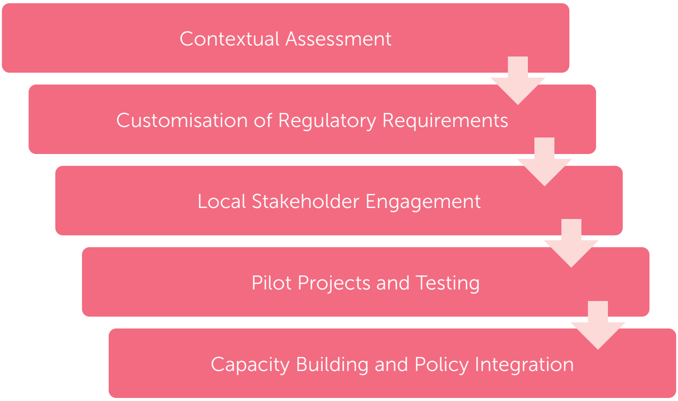
8.1 Key Components of an Inclusive AI Regulatory Framework

An inclusive AI regulatory framework must contain several key components to ensure it is effective, adaptable and fair. First, the framework should be flexible to accommodate different regions' diverse socio-economic, cultural and technological landscapes. This flexibility can be achieved through tiered regulatory requirements considering varying levels of technological development and AI readiness. Second, the framework should prioritise ethical guidelines addressing data privacy, algorithmic bias, and accountability. These guidelines should be developed in consultation with local communities to ensure they reflect cultural values and societal norms. Third, the framework should include robust data governance policies that protect personal data and ensure transparency in AI decision-making processes. Finally, the framework must support capacity-building by investing in education and training programmes to develop a skilled workforce capable of managing and regulating AI technologies (European Commission, 2021; High-Level Expert Group on Artificial Intelligence, 2019).

8.2 Steps for Adapting EU Regulations to Local Contexts

Adapting EU AI regulations to local contexts involves several critical steps, as shown in Figure 4 below.

Figure 4: Steps for Adapting EU Regulations to Local Contexts


Contextual Assessment
Customisation of Regulatory Requirements
Local Stakeholder Engagement
Pilot Projects and Testing
Capacity Building and Policy Integration

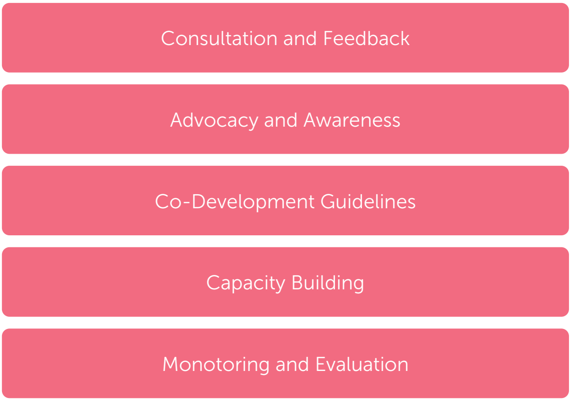
Source: Author (2024)

Adapting EU AI regulations to local contexts involves several critical steps: conduct a thorough assessment of the local socio-economic, cultural and technological environment to identify specific needs and challenges. This assessment should involve gathering input from various stakeholders, including government agencies, industry representatives, academic institutions, and civil society organisations. Modify the existing EU regulations to better fit the local context by creating tiered or phased implementation plans. These modifications might include adjusting compliance requirements based on the region's technological maturity and readiness. Engage local stakeholders in the adaptation process to ensure the regulations are relevant and address the community's concerns. This involves conducting public consultations, workshops, and focus groups to gather feedback and build consensus on the proposed regulations. Implement pilot projects to test the adapted regulations in real-world settings. These projects can provide valuable insights into the practical challenges of enforcement and help refine the regulatory framework before full-scale implementation. Develop and implement training programmes to equip local regulators, industry professionals, and other stakeholders with the necessary skills and knowledge to comply with and enforce the adopted regulations. Ensure that the adapted regulations are integrated with existing national policies and legal frameworks to avoid conflicts and promote coherence in AI governance (South African Government, 2020; Government of Kenya, 2021).

8.3 Role of Local Stakeholders in the Regulatory Process

Local stakeholders play an important role in the regulatory process by providing insights into their communities' specific needs and challenges. Their involvement assures that the regulatory framework is contextually relevant and enjoys broad-based support. Key roles for local stakeholders include those listed in Figure 5 below.

Figure 5: Role of Local Stakeholders in the Regulatory Process


Consultation and Feedback
Advocacy and Awareness
Co-Development Guidelines
Capacity Building
Monotoring and Evaluation

Source: Author (2024)

Participating in consultations and providing feedback on proposed regulations helps identify potential issues and areas for improvement. Raising awareness about the importance of AI regulation and advocating for ethical and fair practices is another key role. Stakeholders can help educate the public and policymakers about the benefits and risks of AI. Collaborating with regulators to develop ethical guidelines and standards that reflect local cultural values and societal norms is essential. This co-development ensures that the guidelines are relevant and widely accepted. It is also important to contribute to training and education initiatives to build local expertise in AI and its regulation. This can involve partnerships with academic institutions, industry, and international organisations. Engaging in the continuous monitoring and evaluation of the regulatory framework assures its effectiveness and adaptability. Stakeholders can provide ongoing feedback and help identify emerging issues that must be addressed (Nigerian Government, 2021).

8.4 Mechanisms for Continuous Evaluation and Improvement of AI Regulations

To ensure that AI regulations remain effective and relevant, it is essential to establish mechanisms for continuous evaluation and improvement. Conducting regular reviews of the regulatory framework to assess its performance and identify areas for improvement is crucial. This can involve periodic assessments, impact evaluations, and stakeholder consultations. Collecting and analysing data on the implementation and outcomes of AI regulations is another key mechanism. This data can provide insights into compliance levels, enforcement mechanisms' effectiveness, and regulations' broader impact on innovation and societal well-being. Setting up formal feedback loops that allow stakeholders to report issues, suggest improvements, and share best practices is also important. This can include public forums, online platforms, and advisory committees. Adopting an adaptive regulatory approach that allows for adjustments based on new evidence, technological advancements, and changing societal needs ensures that regulations can evolve in response to emerging challenges and opportunities. Engaging in international collaboration to learn from the experiences of other regions and incorporate global best practices into the local regulatory framework is essential. This can involve participation in international forums, partnerships with regulatory bodies, and knowledge exchange initiatives (Government of Rwanda, 2019).

8.5 Success Stories of AI Regulation Adaptation in Other Regions

Several regions around the world have successfully adapted AI regulations to fit their unique socio-economic and technological contexts, providing valuable insights for African nations. Singapore has established itself as a leader in AI regulation by developing a comprehensive AI governance framework that balances innovation with ethical considerations. The country's "Model AI Governance Framework" provides detailed guidance on implementing responsible AI systems, emphasising transparency, accountability and fairness. This framework has been well received by both the public and private sectors, fostering an environment conducive to AI innovation while assuring that ethical standards are upheld (Infocomm Media Development Authority, 2020). Canada has adopted a human-centric approach to AI regulation, protecting individual rights and promoting ethical AI development. The Canadian government introduced the "Directive on Automated Decision-Making", which mandates that all federal government departments and agencies use AI in a transparent, fair and accountable manner. The directive includes impact assessments, transparency requirements, and measures to mitigate algorithmic bias, setting a high standard for ethical AI use in government operations (Government of Canada, 2019).

The United Kingdom has implemented a multi-faceted approach to AI regulation, combining legal frameworks with ethical guidelines. The "AI Sector Deal" outlines the government's commitment to fostering AI innovation while addressing ethical and societal concerns. The UK also established the Centre for Data Ethics and Innovation (CDEI) to provide independent advice on AI and data-driven technologies, assuring that ethical considerations are integrated into policy development (UK Government, 2018). These examples illustrate the importance of tailoring AI regulations to local contexts, balancing innovation with ethical considerations, and engaging various stakeholders in the regulatory process. By learning from these success stories, African nations can develop AI regulations that are effective, contextually relevant, and supportive of both innovation and ethical standards.

8.6 Comparative Examples of AI Governance in African and EU Contexts

Comparing AI governance in African and EU contexts reveals commonalities and differences that can inform regulatory adaptation efforts. The EU's AI governance comprises comprehensive regulations emphasising ethical standards, data protection, and human rights. The "Artificial Intelligence Act" categorises AI systems based on risk levels and imposes stringent requirements on high-risk applications, such as those used in critical infrastructure, education, and law enforcement. The EU's holistic approach addresses the entire AI lifecycle from development to deployment and monitoring (European Commission, 2021). In contrast, AI governance in

Africa is still in its early stages, with varying levels of development across different countries. Some nations, like South Africa and Kenya, have made significant progress by developing national AI strategies and regulatory frameworks. South Africa's "National Artificial Intelligence Strategy" focuses on ethical AI development, capacity-building, and promoting AI innovation to drive economic growth. This strategy highlights the importance of building local expertise and infrastructure to support AI technologies (South African Government, 2020).

Similarly, Kenya's "Digital Economy Blueprint" outlines the country's vision for integrating AI into various sectors, emphasising the need for ethical guidelines and regulatory oversight. Kenya's approach aims to harness the potential of AI to transform sectors such as agriculture, healthcare and education while ensuring that ethical considerations are integrated into AI governance (Government of Kenya, 2021). These comparative examples illustrate the stages of AI governance between the EU and African contexts. The EU's mature regulatory framework provides a comprehensive and holistic governance model, while African countries are making strides in developing their own strategies that reflect their unique socio-economic and technological landscapes. By understanding these commonalities and differences, African nations can adapt and refine their AI governance frameworks to ensure they are effective and contextually relevant.

9 Lessons Learned from These Case Studies

Several important lessons can be drawn from these case studies to inform the development of AI regulations in Africa. Successful AI governance frameworks are holistic, addressing the entire AI lifecycle and incorporating ethical, legal and societal considerations. Flexibility is crucial to accommodate different regions' diverse socio-economic and technological contexts. African countries should develop adaptable regulatory frameworks that can evolve with technological advancements and changing societal needs. Engaging various stakeholders, including government agencies, industry representatives, academia, and civil society, is essential for developing effective AI regulations. Inclusive consultation processes ensure that diverse perspectives are considered, leading to more robust and contextually relevant regulatory frameworks. Investing in education and training programmes to build local expertise in AI is critical for effective regulation and innovation. Partnerships with international organisations and universities can support capacity-building efforts, ensuring regulators and industry professionals have the necessary skills and knowledge. Establishing clear ethical guidelines and standards is fundamental to fostering public trust and ensuring the responsible use of AI. These guidelines should reflect local cultural values and societal norms, addressing data privacy, algorithmic bias, and accountability. Implementing mechanisms for continuously evaluating and improving AI regulations is vital for keeping pace with technological advancements and emerging challenges. Regular reviews, data collection, and stakeholder feedback can help refine regulatory frameworks and ensure their ongoing relevance and effectiveness. Collaborating with other regions and learning from their experiences can enhance African AI governance. Participation in international forums, partnerships with regulatory bodies, and knowledge exchange initiatives can provide valuable insights and best practices that inform local regulatory efforts. By incorporating these lessons, African nations can develop effective and inclusive AI regulatory frameworks that promote innovation, protect individual rights, and safeguard the equitable distribution of AI benefits

10 Recommendations

Policymakers should thoroughly assess their local socio-economic, cultural and technological environments to customise AI regulations effectively. This involves gathering input from various stakeholders and understanding each region's unique challenges and opportunities. Creating tiered or phased regulatory requirements based on a country's technological development and AI readiness is essential. This flexibility ensures that regulations are relevant and achievable for different regions. Developing clear ethical guidelines and standards that reflect local cultural values and societal norms is also crucial. These guidelines should encompass data privacy, algorithmic bias, accountability, and transparency in AI systems. Engaging a broad range of stakeholders, including government agencies, industry representatives, academia, and civil society, in developing AI regulations is vital. Inclusive consultation processes, such as public consultations, workshops, and focus groups, can help gather diverse perspectives and build consensus around the proposed regulations. Conducting public awareness campaigns to educate citizens about the benefits and risks of AI technologies and the importance of ethical AI governance can help build public trust and support for AI regulations. Fostering partnerships between local and international organisations to support capacity-building initiatives, research collaborations, and sharing best practices can add to the effectiveness of AI governance efforts. Governments and research institutions should invest in studies exploring AI's socio-economic impacts, the effectiveness of different regulatory approaches, and the ethical implications of AI technologies. This research can inform policy development and regulatory adjustments.

Encouraging participation in international forums, working groups and partnerships focusing on AI governance can facilitate the harmonisation of AI regulations and address global challenges more effectively. Establishing mechanisms for continuously evaluating and improving AI regulations is also essential. Regular reviews, data collection and stakeholder feedback can help refine regulatory frameworks and ensure their ongoing relevance and effectiveness. By implementing these recommendations, policymakers and regulators can develop inclusive and adaptable AI regulatory frameworks that promote innovation, protect individual rights, and provide equitable AI benefits across different regions.

11 Conclusion

This chapter has explored the complexities and challenges of adapting the EU AI regulations to African contexts, comprehensively analysing both regions' regulatory landscapes. Key findings include the significant socio-economic, cultural and technological differences between the EU and African nations, calling for a flexible and tailored approach to AI regulation. The study highlighted African countries' unique challenges, such as resource limitations, infrastructural disparities, and varying levels of AI readiness. Addressing ethical concerns and ensuring equitable benefits from AI technologies was also emphasised. The presented successful examples of AI regulation adaptation from other regions, such as Singapore and Canada, provide valuable lessons to inform the development of inclusive and effective AI regulatory frameworks in Africa. The international collaboration proved to be a critical factor in developing and implementing effective AI regulations. By engaging in global dialogues, sharing best practices, and fostering partnerships, regions can benefit from collective knowledge and experience. International collaboration can also facilitate harmonising AI governance, promoting consistency in ethical standards and regulatory approaches across different regions. This collaborative effort is essential for addressing the global challenges associated with AI, such as data privacy, algorithmic bias, and the ethical use of AI technologies. Creating a balanced and effective AI regulatory environment requires a holistic approach that integrates ethical, legal and societal considerations. Regulations must be adaptable to cater to the specific needs and contexts of different regions, ensuring they promote innovation while protecting individual rights and societal values. Engaging local stakeholders in the regulatory process, building capacity via education and training, and establishing robust mechanisms for continuous evaluation are crucial steps for achieving this balance. Ultimately, the goal is to develop inclusive, transparent, and responsive AI governance frameworks for the evolving landscape of AI technologies.

REFERENCES

- Agrawal, A., Gans, J., & Goldfarb, A. (2018). Prediction machines: the simple economics of artificial intelligence. Harvard Business Review Press.
- Ding, J., Triolo, P., & Sacks, S. (2018). Chinese Interests Take a Big Seat at the AI Governance Table. New America, 20.
- European Commission. (2018). Draft Ethics Guidelines for Trustworthy AI. Available online: https://digital-strategy.ec.europa.eu/en/library/draft-ethics-guidelines-trustworthy-ai
- European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Available Online: https://resourcecenter.cis.ieee.org/government/eu/cisgovph0040
- European Commission. (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Available online: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence
- Executive Office of the President. (2020). Guidance for Regulation of Artificial Intelligence Applications. Office of Management and Budget, USA. Available online: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf
- Germany Federal Ministry for Economic Affairs and Climate Action. (2018). Artificial Intelligence Strategy. Available online: https://www.de.digital/DIGITAL/Redaktion/EN/Meldungen/2018/2018-11-16-federal-government-adopts-artificial-intelligence-strategy.html
- Government of Canada. (2019). Directive on Automated Decision-Making. Available online: https://www.tbs-sct.canada.ca/pol/doc-eng.aspx?id=32592
- Government of France. (2018). AI for Humanity: French Strategy for Artificial Intelligence. Available online: https://www.aiforhumanity.fr/en/
- Government of Kenya. (2021). Kenya's Digital Economy Blueprint. Available online: https://cms.icta.go.ke/sites/default/files/2022-04/Kenya%20Digital%20Masterplan%202022-2032%20Online%20Version.pdf
- Government of Rwanda. (2019). Rwanda's Vision 2050. Available Online: https://www.minecofin.gov.rw/fileadmin/user_upload/Minecofin/Publications/REPORTS/National_Development_Planning_and_Research/Vision_2050/English-Vision_2050_Abridged_version_WEB_Final.pdf
- High-Level Expert Group on Artificial Intelligence. (2019). Ethics Guidelines for Trustworthy AI. Available online: https://merlin.obs.coe.int/article/8608
- Mhlanga, D. (2023). Artificial intelligence and machine learning for energy consumption and production in emerging markets: a review. Energies, 16(2), 745.
- Mhlanga, D. (2024). Generative AI for Emerging Researchers: The Promises, Ethics, and Risks. Ethics, and Risks (February 24, 2024).

- Mhlanga, D., & Salih, S. (2023). After Being Left Out of the First, Second, and Third Industrial Revolutions, Is Africa Finally Prepared for the Fourth Industrial Revolution? In The fourth industrial revolution in Africa: Exploring the development implications of smart technologies in Africa (pp. 35-51). Cham: Springer Nature Switzerland.
- Ministry of Economic Affairs and Employment of Finland. (2017). Finland's Age of Artificial Intelligence. Available online: https://www.aepia.org/wp-content/uploads/2022/11/TEMrap_47_2017_verkkojulkaisu.pdf
- Nigerian Government. (2021). Nigeria's AI Policy Framework. Available online: https://paradigmhq.org/wp-content/uploads/2021/11/Towards-A-Rights-Respecting-Artificial-Intelligence-Policy-for-Nigeria.pdf
- Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., ... & Bengio, Y. (2019). Tackling climate change with machine learning. arXiv preprint arXiv:1906.05433.
- Salih, S., & Mhlanga, D. (2023). Blockchain for Food Supply Chain: Trust, Traceability, and Transparency Enhancement, How Can Africa Benefit? The Fourth Industrial Revolution in Africa: Exploring the Development Implications of Smart Technologies in Africa, 311-325.
- Singapore Personal Data Protection Commission. (2020). Model AI Governance Framework. Available online: https://www.pdpc.gov.sg/help-and-resources/2020/01/model-ai-governance-framework
- South African Government. (2023). National Artificial Intelligence Strategy. Available Online: https://www.dcdt.gov.za/images/phocadownload/AI_Government_Summit/National_AI_Government_Summit_Discussion_Document.pdf
- Topol, E. (2019). High-performance medicine: the convergence of human and artificial intelligence. Nature Medicine, 25(1), 44-56.
- UK Government. (2018). AI Sector Deal. Available online: https://assets.publishing.service.gov.uk/media/5d31f49ce5274a14ec200c0c/AI_Sector_Deal_One_Year_On__Web_.pdf

Chapter 10

The AI Act Balancing Innovation and Regulation in Welfare: A Case Study of the e-Welfare System in Slovenia

Dejan Ravšelj

Matej Babšek

1 Introduction

The rapidly developing systems of artificial intelligence (AI) hold unimagined potential to fundamentally alter existing social and economic structures and become one of the most important technological breakthroughs in modern times for individuals, businesses and governments (Misuraca & Van Noordt, 2020).

As a fundamental societal subsystem, national welfare systems are also not immune to these influences and thus the adoption of AI in their own processes and services should consider fundamental values and broader systemic aspects of the national and European space alongside technological aspects, leading to higher quality and more efficient, trustworthy and user-centred public services for citizens (Henman, 2020). The main objectives of the European Union (EU) are to become a world leader in the development and deployment of remarkable, ethical and safe AI and to promote a human-centred approach on a global scale (Misuraca & Van Noordt, 2020). In this context, the Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence

(AI Act) is a ground-breaking legal framework aimed at fostering innovation while ensuring the safe and ethical use of AI technologies. This is particularly important in the field of social welfare, which deals with fundamental aspects of human dignity and human rights pursued in administrative procedures at the centre of public welfare systems. While on one hand, AI-powered systems in this area streamline and improve the delivery of public services, on the other these systems have also sparked considerable controversy as they have the potential to perpetuate prejudice and undermine individual rights. Systems such as SyRI in the Netherlands, which was developed to detect child benefit fraud, or the MiDAS project in the United States, which detected fraud in cash transfers for the unemployed, show that the misuse of AI algorithms can lead to discrimination and tragic life situations for tens of thousands of individuals and families. These cases show the risks and challenges associated with AI-supported decision-making, such as invasion of privacy, discrimination and lack of transparency.

This paper therefore first provides an overview of the legal framework created by the AI Act and its impact on welfare procedures. It then outlines the characteristics of AI-supported automated decision-making in such procedures. This is followed by a presentation of the most significant abuses of AI in these procedures, as a reminder that technological solutions without adequate legal and social oversight do not fulfil their purpose in the pursuit of shared societal values. What follows next is a comprehensive analysis of the Slovenian AI-powered e-Welfare system introduced in 2012, which pursues the goals of a fast, efficient, fair and transparent assessment of social rights in Slovenia. This system is also analysed through a comparative analysis of similar systems in this field in other EU countries. The final section of this chapter summarises the main findings of the previous sections and presents conclusions and recommendations for decision- and policy-makers on the EU and national levels.

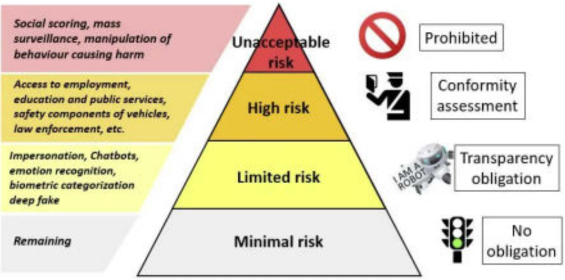
2 The AI Act for Welfare

The EU AI Act, the world’s first comprehensive AI law, seeks to regulate the use of AI to ensure better conditions for its development and utilisation. As part of the EU’s digital strategy, the Parliament adopted the AI Act in March 2024, followed by it being approved by the Council in May 2024 and in force since August 2024. Originally proposed by the European Commission in April 2021 and politically agreed in December 2023, the AI Act creates a common framework for AI systems in the EU. It classifies AI systems based on a risk-based approach and sets out various requirements and obligations to effectively manage their use and deployment (European Parliament, 2024).

For the purposes of the AI Act, an "AI system" means a machine-based system designed to operate with varying levels of autonomy and which 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 (Article 3(1)). AI exhibits human characteristics like thinking, learning, planning and creativity. AI enables a technical system to sense the environment, process the collected or accessible data, and solve a problem according to a specific goal and 'learn' by adapting its way of working based on input information, algorithms and inbuilt criteria and intermediate results. The use of AI can support socially and environmentally beneficial outcomes and provide companies in the European economy with decisive competitive advantages by improving forecasting, optimising operations and resource allocation, and providing tailored, appropriate value-added services (European Commission, 2021).

The EU AI Act is an important legal framework that aims to establish harmonised rules for the development, commercialisation and use of AI in the EU. The main added value of the AI Act, apart from its immediate applicability in a harmonised form across the whole EU, is the introduction of a system of different levels of risk, which is extremely important from the perspective of welfare or public intervention in general. The Act takes a pyramid, risk-based approach to regulating AI systems by categorising them on four levels of risk: unacceptable risk, high risk, limited risk and minimal risk (Edwards, 2021), as shown in Figure 1. These levels further lead to prohibited, permitted limited or relatively open uses of AI, with the European Commission identifying areas where one risk level or the other applies. The AI Act seeks to harmonise the different approaches in the EU so that AI tools can be used in the future with as few detours as possible.

Figure 1: Risk levels according to the EU AI Act


Social scoring, mass surveillance, manipulation of behaviour causing harm
Unacceptable risk
Prohibited
Access to employment, education and public services, safety components of vehicles, law enforcement, etc.
High risk
Conformity assessment
Impersonation, Chatbots, emotion recognition, biometric categorization deep fake
Limited risk
I AM A ROBOT
Transparency obligation
Remaining
Minimal risk
No obligation

Source: Telefónica, 2024

According to the AI Act, the risk assessment also takes other aspects into account where the potentially harmed or affected persons are in a vulnerable position compared to the user of the AI system, in particular due to an imbalance of power, knowledge, economic or social circumstances or age. For example, procedures in the area of welfare, together with the areas of education, employment, justice, home affairs etc., are predominantly categorised on the high-risk level (European Commission, 2021). This requires an ex-ante and ex-post conformity assessment based on the impact and specificities of each area. High-risk AI systems are subject to strict legal requirements due to their potential impact on fundamental rights and safety. The conformity assessment framework aims to mitigate the risks associated with high-risk AI systems and ensure they operate within acceptable safety and ethical standards to promote trust and safety in AI applications in the EU (Edwards, 2021).

In addition, the AI Act distinguishes providers who develop or own an AI system in order to bring it to market or use it in their own name from AI deployers (providers and deployers) since the role of the decision-making authority in the welfare system changes in this respect. Given the potential dangers posed by the use of AI in welfare systems, the AI Act focuses on ensuring transparency and human oversight to make these high-risk AI systems more transparent. These goals could be supported by the concept of explainable AI (XAI), which helps users (i.e., welfare officials making decisions about citizens' social rights) of AI- systems to work more effectively by gaining a clearer understanding of how these systems function (Panigutti et al., 2023).

> ”
>
> High-risk AI systems are subject to strict legal requirements due to their potential impact on fundamental rights and safety.

Certain aspects of the AI Act are especially important for AI-supported procedures in welfare, e.g.:

1. **Human-centred AI:** the AI Act emphasises the development of AI systems that serve people and respect human dignity and personal autonomy (AI Act, Recital 6). This is crucial in welfare procedures where a personalised and empathetic service is required to provide effective support to beneficiaries.
2. **Risk-based approach:** the AI Act takes a risk-based approach and categorises AI systems based on their potential impact on health, safety and fundamental rights (AI Act, Recital 26). Strict requirements apply to high-risk AI systems used in welfare services, such as eligibility verification or benefit distribution systems, to make sure they do not cause harm or discrimination.
3. **Transparency and accountability:** the AI Act prescribes transparency in AI operations and requires that AI systems must be explainable and their decisions comprehensible (AI Act, Recital 27). This is particularly important in welfare procedures to maintain trust and allow beneficiaries to understand how decisions that affect their lives are made.
4. **Data protection and privacy:** the AI Act ensures that AI systems comply with applicable data protection laws and that the personal data of welfare recipients is protected (AI Act, Recital 10). Given the sensitive nature of data related to welfare benefits, this compliance is crucial for protecting the privacy of individuals and preventing the misuse of their information.
5. **Prevention of harmful practices:** certain AI practices that could lead to manipulation or exploitation are prohibited under the AI Act (AI Act, Recital 29). This protection is essential in the context of welfare procedures to stop AI from being used in a way that could unfairly influence or harm vulnerable people.

3 AI-Supported Automated Decision-Making in Welfare Procedures

Digital transformation in the welfare sector encompasses a range of mechanisms, from simple electronic communication to advanced AI technology solutions. These solutions are used both internally, between public authorities and public servants, and externally, between public authorities and citizens (Misuraca et al., 2020). The use of machine learning for automated administrative decisions in the welfare sector allows decisions about and for citizens to be made based on automated data processing. This data can be original or come from secondary analyses, profiling etc. In these procedures, procedural rules support the use of new AI solutions by assuring basic procedural safeguards, thereby promoting more up-to-date, evidence-based and democratically accountable decisions by public authorities (Coglianese, 2021).

For automated decision-making to be effective, it should be considered not only in terms of the technical infrastructure and its processes, but also as the creation of meanings about what technology is and what it can do. This includes the creation of a value-based socio-technical imaginary about AI (Kaun, 2022). Such a perspective is essential for welfare procedures that stress practical aspects and interactions between authorities and citizens (Ranchordas, 2022). While introducing AI technologies into welfare processes and systems, it is important to consider the moral/ethical issues and philosophical dilemmas that may be overlooked (Larsson & Haldar, 2021). This means the practice of automated decision-making should be guided by the principles of good governance and incorporate empathy into automated administrative decision-making to enhance human interaction and promote the well-being and prosperity of individuals (Coglianese, 2021). In decision-making processes to assess eligibility and the scope of social benefit entitlements, public authorities intervene in the legal positions of citizens and other parties in the public interest following the principles of equity, redistribution and accessibility. These relationships are enshrined in law and innovation is restricted. Further, inadequate technological systems without values, rules and relationships between stakeholders can lead to technocratism rather than improvements when AI is introduced (Ranchordas, 2022).

Welfare procedures guarantee rights within the framework of the constitutional principle of the Democratic and Welfare state as a fundamental human right intended for those who find themselves in social need for objective reasons (Babšek & Kovač, 2023b). Unlike other administrative procedures, welfare

> ”
>
> Substantive legal rights in the field of social welfare are only effectively acquired if the procedure to enforce them is regulated by law and adequately supported to ensure equality and adjustments for citizens.

procedures should be less formal and simpler for parties unfamiliar with the law. However, theoretical considerations (Ranchordas, 2022) suggest that the administrative procedures for enforcing these rights in Western welfare states are among the most administratively rigid in the world. This rigidity is partly due to the unique welfare state arrangements and associated public interests of individual European states, which prevent convergence (Kuhlmann & Wollmann, 2019). In cases where welfare rights are at stake, the public interest often coincides with individual rights as authorities endeavour to create equal opportunities through positive discrimination and social support. The importance of procedures in achieving broader socially accepted goals is emphasised by international and constitutional guarantees and represents a common point of development in the EU and beyond (Auby, 2014). The welfare procedure should therefore fulfil its objective of realising substantive legal rights and serve as both a mechanism and an objective of the welfare system in its own right. Substantive legal rights in the field of social welfare are only effectively acquired if the procedure to enforce them is regulated by law and adequately supported to ensure equality and adjustments for citizens, regardless of whether the need is permanent or temporary. If the process is not predictable and user-friendly, it can hinder legal rights, especially for vulnerable groups who lack knowledge or resources (Ranchordas, 2022). Therefore, under the guise of de-bureaucratisation and digitalisation, legal regulations that disadvantage the most vulnerable citizens should not be created.

In the context of social welfare procedures, the integration of AI is inextricably linked to the use of Big Data, which significantly adds to the efficiency and accuracy of social welfare administration. AI utilises the volume, velocity, variety, veracity and value of data to improve decision-making and service delivery (Dwivedi et al., 2021; Pencheva et al., 2020). The term “volume” refers to the processing of large data sets that ensure comprehensive data analysis; “velocity” refers to the speed of data processing that enables rapid response to applications. “Variety” encompasses different types of data, from structured numerical input to unstructured text, assuring comprehensive analysis. “Veracity” ensures the reliability of data by cross-referencing information from different sources, reducing errors and fraud. “Value” emphasises the benefits of streamlined processes and better resource allocation (ibid.). With advances in emerging technologies, AI combined with Big Data can further revolutionise welfare systems. These technologies enable efficient text analysis, document review and accessible communication channels, making welfare procedures more effective, inclusive and equitable (Wang et al., 2021).

In the field of welfare, AI can be applied on two different levels: artificial narrow intelligence (ANI) and artificial general intelligence (AGI) (Sun & Medaglia, 2019). Narrow AI, also known as weak AI, is designed to handle specific or limited tasks such as speech recognition and recommendation systems. It uses machine learning algorithms to process large amounts of data, recognise patterns and make predictions with high speed and accuracy, yet it cannot generalise or understand relationships beyond the intended functions (Kaplan & Haenlein, 2019). On the other hand, general AI aims to emulate the cognitive abilities of humans and perform intellectual tasks such as reasoning, learning and autonomous problem-solving, which brings with it numerous ethical and societal challenges (Sun & Medaglia, 2019). In addition, there is the theoretical concept of super AI (ASI – artificial super intelligence), which would surpass human intelligence and solve problems that go beyond human capabilities, albeit it is hypothetical and currently unrealisable (Kaplan & Haenlein, 2019).

4 Cases of AI Misuse in Welfare

AI holds significant promise for enhancing the efficiency and effectiveness of welfare systems. Social security institutions are increasingly leveraging AI to enable more proactive and automated delivery of social services. Although the full potential of these technologies has yet to be completely tested and explored, they are already yielding positive outcomes in crucial areas. AI is helping address issues like errors, evasion and fraud while also developing effective approaches and automated solutions to address customers’ concerns. This ultimately improves social services, including social care, communication with insured individuals, and the management of welfare benefits. Still, the successful implementation of AI solutions entails a variety of complex challenges that could hinder their potential for positive impact.

Without careful design, monitoring and refinement in accordance with key social policy principles such as equal rights and social justice, AI systems may produce significant negative effects for individuals, organisations and societies. These effects could worsen existing social problems by increasing inequality and discrimination, casting doubt on a government's capability to adequately protect and serve the citizens (Ananny, 2016; Hassan, 2022).

Several scandals have recently shaken various fields across the EU, underscoring growing concerns about the ethical implications of AI and automated systems in public services. Outlined below are some of the most notable scandals across different sectors: in the Netherlands, SyRI (System Risk Indication) affected social welfare and public administration, while the Toeslagenaffaire (Childcare Benefits Scandal) impacted taxation and childcare benefits; in Austria, the AMS (Austrian Public Employment Service) algorithm scandal involved employment services and labour market classification; and in Spain, VioGén (Gender Violence Prevention System) concerned public safety and gender-based violence prevention.

The SyRI (System Risk Indication) scandal in the Netherlands involved an automated system used by the Dutch government to detect welfare fraud by analysing data from various public sources, including social security, employment and housing records. The system was intended to identify individuals or families at risk of committing fraud, but it faced significant backlash for disproportionately targeting low-income neighbourhoods, often populated by ethnic minorities. Critics argued that SyRI's lack of transparency and the potential for discrimination violated individuals' rights. In February 2020, a Dutch court ruled that SyRI had breached the European Convention on Human Rights, particularly the right to privacy, and ordered the system to be suspended. This ruling was a landmark decision, spotlighting the need for transparency, fairness and protection against discrimination in the use of AI and data-driven tools by governments (Algorithm Watch, 2020a; Van Bekkum & Borgesius, 2021).

The Toeslagenaffaire (Childcare Benefits Scandal) was a major political crisis in the Netherlands, where thousands of families were wrongfully accused of fraudulently claiming childcare benefits. The Dutch Tax and Customs Administration's automated system had identified supposed fraud cases, bringing devastating consequences for those falsely accused. Many families, especially those with dual nationalities and from ethnic minority backgrounds, were forced to repay large sums they did not owe, causing severe financial distress, the loss of homes, and social stigmatisation. The scandal exposed deep flaws in the automated system and its lack of human oversight, culminating in the resignation of the entire Dutch government in January 2021. The affair sparked widespread outrage and discussions concerning the ethical use of AI in public administration, highlighting the dangers of discrimination and the critical need for transparency and accountability in government systems (Amnesty International, 2021; Politico, 2022).

The AMS (Austrian Public Employment Service) algorithm scandal revolved around a controversial AI-driven system introduced by Austria's AMS to classify job seekers based on their predicted chances of reemployment. The algorithm used data-driven models to place individuals in three categories, which then determined the level of support and resources they received, such as training and educational opportunities. The system faced a serious backlash for its potential to reinforce and exacerbate existing social inequalities, particularly affecting women, older individuals, and immigrants, who were often assigned lower employability scores. Critics condemned the algorithm for its opaque decision-making process and lack of transparency generally, which made it difficult for those affected to understand or challenge their classifications. In 2020, following a public outcry and scrutiny from the Austrian Data Protection Authority, use of the AMS Algorithm was suspended. The authority raised concerns about its compliance with the General Data Protection Regulation (GDPR) and the risk of discriminatory outcomes. The AMS algorithm scandal shows significant ethical issues are related to the use of AI in public services, notably the need for greater transparency, fairness, and human oversight in automated decision-making processes (Digital Watch, 2019; Sekwenz, 2022).

> ”
>
> The AMS algorithm scandal shows significant ethical issues are related to the use of AI in public services.

The VioGén (Gender Violence Prevention System) in Spain was designed as an automated system to help prevent domestic and gender-based violence by assessing the risk levels of potential victims and coordinating protective measures. While the system aimed to improve the safety of vulnerable individuals, it became controversial due to its reliance on algorithms that sometimes produced inaccurate risk assessments. The VioGén system, intended to flag high-risk cases for immediate intervention, occasionally failed to identify real threats or, conversely, flagged low-risk cases as high-risk, leading to misallocation of resources. These inaccuracies held serious consequences, including instances where individuals at genuine risk did not receive the necessary protection, resulting in tragic outcomes. The controversy around VioGén exposed the limitations of relying heavily on automated decision-making in critical areas like public safety, revealing the need for more nuanced human oversight, improved algorithmic accuracy, and greater transparency in how risk assessments are conducted. The scandal sparked debates about the ethical use of AI in protecting vulnerable populations and the importance of ensuring that such systems are both reliable and fair (Algorithm Watch, 2020b; Eticas, 2022).

5 Case Study of the e-Welfare System in Slovenia

In 2012, Slovenia introduced e-Welfare (Slov. e-Sociala), an innovative digital system within its social welfare system that primarily uses AI to simplify administrative decision-making in the field of social welfare. This comprehensive system is embedded in the information system of Slovenian social work centres and aims to reduce bureaucracy and significantly influence the distribution of welfare benefits. The e-Welfare system uses networks of algorithms and machine learning to process data and make decisions on citizens' applications for welfare benefits. This automation ensures that decisions are made based on objective, evidence-based data, increasing transparency and accountability within the welfare administration. Through the use of AI, the system aims to ensure the efficient, fair and transparent redistribution of public funds while upholding the constitutional principles of the rule of law and the welfare state to protect vulnerable groups from arbitrary action by the authorities. The e-Welfare system has shown that it has the potential to improve the speed and objectivity of decisions, but has also raised concerns about possible abuses, data protection issues and the need for an appropriate legal framework for the use of AI in social welfare (Babšek & Kovač, 2023a).

5.1 Practical SWOT Analysis

The introduction of the e-Welfare system, which is based on AI-supported decision-making in the field of Slovenian welfare law, is an important step towards Slovenia's digital transformation. This system aims to improve efficiency, fairness and transparency in the redistribution of welfare benefits while upholding constitutional principles and protecting vulnerable groups. Table 1 shows a SWOT analysis assessing the strengths, weaknesses, opportunities and threats of this system based on data from interviews, focus groups and empirical studies on the e-Welfare system (see Babšek & Kovač, 2023a; Babšek & Kovač, 2023b for more details).

Table 1: SWOT analysis of the e-Welfare system in Slovenia

Strengths:	**Weaknesses:**
• **Efficiency and speed:** e-Welfare system automates repetitive and time-consuming administrative tasks, resulting in faster decision-making processes and faster service delivery • **Objectivity and fairness:** By using AI algorithms, decisions are based on objective data analysis, reducing human bias and ensuring equal treatment of all applicants • **Cost efficiency:** e- Welfare automation reduces the need for extensive human resources, resulting in cost savings for the administration • **Improved data processing:** AI enables fast and accurate processing of large amounts of data, improving the overall redistribution of public funds	• **Lack of personal attention:** e-Welfare lacks the empathy and understanding that human interactions provide, potentially penalising vulnerable people who need personal support • **Complexity of implementation:** The integration of AI-powered e-Welfare into existing social services is complex and requires significant adjustments to legal norms and organisational processes • **Risk of errors:** If e-Welfare is not properly managed and updated, significant errors can occur, as the problems with the informational calculation of child benefit in Slovenia have already shown • **Data protection concerns:** The handling of large amounts of sensitive personal data raises data protection and privacy concerns

Opportunities:	Threats:
• **Scalability and innovation:** Lessons learned from the e-Welfare system can serve as a model for other sectors and regions and encourage wider adoption of AI technologies in service delivery • **Improving service delivery:** The efficiencies and data insights gained through AI can lead to more informed policymaking and better targeted welfare services • **Improved public trust:** Transparent and objective decision-making processes can add to public trust in government institutions and their ability to effectively manage welfare programmes • **Proactive support:** AI can help identify individuals who may not be able to assert their rights themselves so that social work centres can provide proactive support	• **Technological dependence:** Over-reliance on e-Welfare can lead to significant disruption when technical problems arise and there is a lack of personal interaction • **Ethical and legal challenges:** The use of AI in decision-making raises ethical and legal issues, particularly in relation to the accountability and transparency of authorities in automated decision-making • **Resistance to change:** Stakeholders, including employees and beneficiaries, may resist the transition to automated systems, leading to implementation issues • **Potential for abuse:** Without adequate safeguards, there is a risk that AI systems will be used for purposes other than those intended, e.g., for discriminatory practices

Sources: own elaboration, based on Babšek & Kovač, 2023a; Babšek & Kovač, 2023b

Implementation of the e-Welfare system has faced considerable challenges, in particular a major error in 2021 that affected the assessment of child benefits following a technical error in obtaining data on beneficiaries' income from the Slovenian tax administration, which led to underpayment for about 20% of beneficiaries (Babšek & Kovač, 2023a). As Slovenian NGOs emphasise, the lack of personal support is also a major weakness of the e-Welfare system. Vulnerable people who need personalised support may find it difficult to navigate the system without human interaction (ibid.). In order to address the ethical and legal challenges of e-Welfare, it is imperative that clear guidelines and safeguards are established. Ensuring accountability and transparency in automated decision-making is essential for preventing abuse and maintaining public trust. The integration of AI into the e-Welfare system should strike a balance between efficiency and the protection of individual rights and privacy (Babšek & Kovač, 2023b).

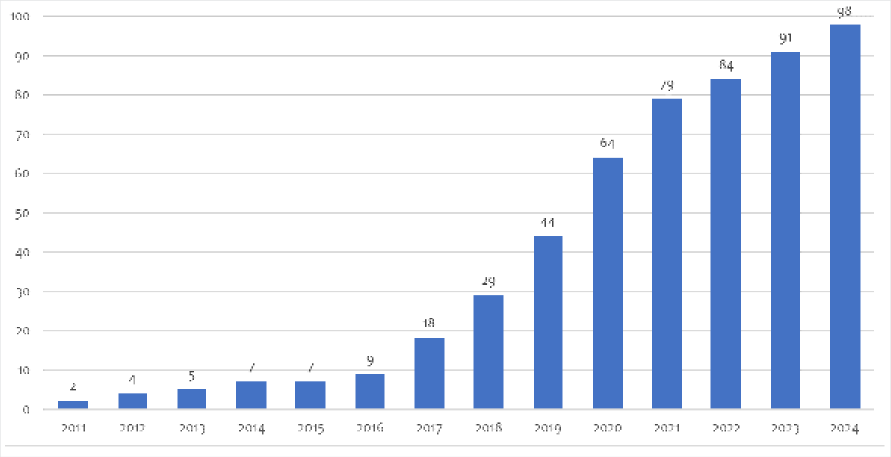
To maximise the benefits of the e-Welfare system in Slovenia and overcome its challenges, several opportunities for improvement have been identified:

1. **Strengthened legal framework:** closer alignment of the e-Welfare system with existing laws will help mitigate legal issues and assure compliance with legal standards.
2. **Robust technical infrastructure:** investment in a robust technical infrastructure and continuous system monitoring will help to avoid technical failures and ensure the system's smooth operation.
3. **Personalised support:** combining AI with human interaction can provide a more holistic approach to social welfare and make sure that vulnerable people receive the individual support they need. AI can take over routine tasks, allowing social workers to focus on more complex cases that require individualised support.
4. **Ongoing training and support:** ongoing training and support for staff carrying out procedures in the e-Welfare system will help them use the system effectively and resolve any issues that arise. This includes training on both the technical aspects of the system and the importance of empathy in service delivery.
5. **Improved data analytics:** the use of advanced data analytics can improve the system's ability to predict and resolve emerging social issues. By analysing trends and patterns in welfare data, the system can help policymakers develop more effective policies and allocate resources more efficiently.

5.2 Comparative Analysis with Similar Applications in the EU

The latest evidence indicates that 98 AI use cases have been recorded in the field of social protection within the EU public sector (Figure 2). The adoption of AI began modestly with just two cases in 2011, but has since steadily accelerated, particularly from 2017 onwards. By 2018, the number of use cases had risen to 29, more than doubling the count from the previous year. This upward trajectory continued, with substantial annual increases, culminating in 98 use cases by 2024. This cumulative growth underscores the progressive adoption of AI-driven initiatives aimed at enhancing social protection mechanisms across EU member states. Key advancements in AI during this period significantly contributed to its increased use in social protection within the EU public sector from 2017 onwards. Namely, 2017 was pivotal due to the convergence of several factors that accelerated AI innovation and adoption. First, the proliferation of connected devices, driven by the rise of sensors and the Internet of Things (IoT), generated vast amounts of data for AI systems, boosting their analytical capabilities. Concurrently, reduced computing costs, propelled by Moore's Law, made AI more scalable and accessible, facilitating its deployment across various public sector applications, including social protection. The explosion of data availability, driven by digitalisation, social media, and smartphones, also provided the essential raw material for training AI models, enabling them to function more effectively in public sector contexts. Lastly, advancements in machine learning, especially deep learning, significantly improved AI's ability to recognise patterns, make decisions and predict outcomes, further enhancing its utility in public sector services such as healthcare and social welfare (World Economic Forum, 2017).

Figure 2: Cumulative number of AI use cases in social protection across the EU public sector

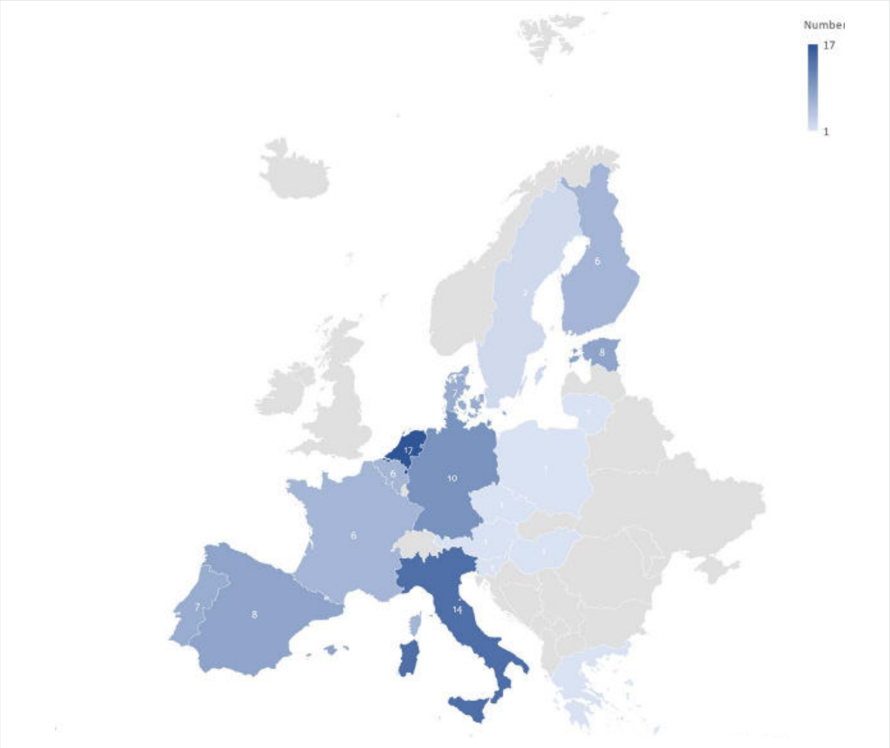


Source: European Commission, 2024

Most of the AI use cases in social protection are on the national level (49 cases), followed by local initiatives (34 cases), regional efforts (12 cases), and a smaller number across multiple countries (3 cases). This distribution suggests that AI technologies are primarily being deployed on the national level where they can impact large populations and address broad social protection needs. National governments may have more resources and authority to implement wide-scale AI solutions, such as fraud detection, public employment services, and family allowance systems, which require extensive data and infrastructure. Local initiatives, though fewer in number, play a crucial role in addressing specific community needs, leveraging AI to tailor services to local populations. These might include localised public health programmes, community-based social services, or municipal management systems that benefit from AI-driven efficiencies. Regional efforts are less common, possibly due to the complexity of coordinating AI initiatives across multiple jurisdictions within a country, each with its own regulatory environment and needs. These efforts might focus on cross-border regions or specific areas where regional cooperation is essential, such as managing shared natural resources or regional economic development. Finally, the few use cases that span multiple countries highlight the potential for AI to address transnational issues, such as migration, international social security agreements, or cross-border public health challenges. These projects require significant collaboration and alignment between different countries' policies and data systems, which can be challenging but offer valuable opportunities for addressing global social protection challenges (Ohlenburg, 2020).

Geographically, the distribution of AI use cases in social protection across EU member states is uneven, with a noticeable concentration in Western Europe (Figure 3). The Netherlands leads with 17 cases, followed by Italy with 14 and Germany with 10. Countries like Estonia and Spain also show a relatively high number of use cases, with 8 each, while Denmark and Portugal have 7. Belgium, Finland and France each report 6 cases. In contrast, Sweden has only 2 cases, and several countries, including Austria, Czechia, Greece, Hungary, Lithuania, Poland and Slovenia, each have just 1. This distribution reveals that some EU member states are considerably more advanced in integrating AI into social protection systems than others.

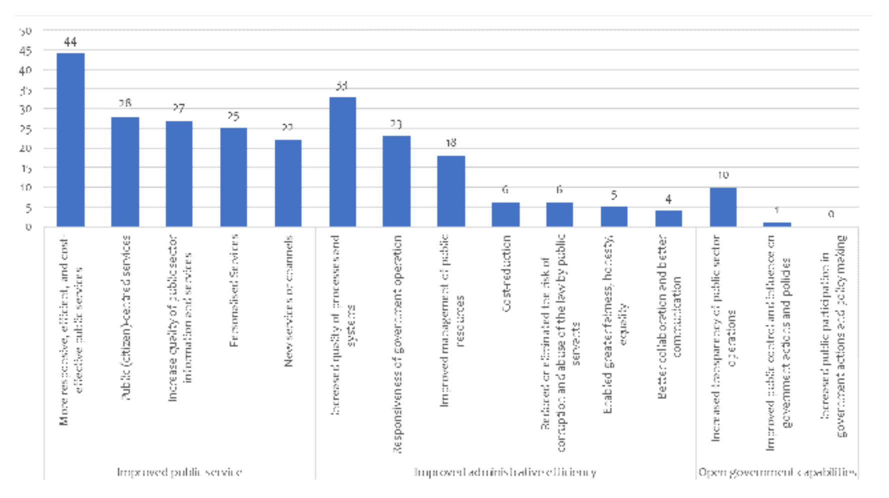
Figure 3: Number of AI use cases in social protection across the EU member states



Source: European Commission, 2024

Governments are increasingly harnessing AI to craft more effective policies, make informed decisions, enhance communication and engagement with citizens, and significantly improve the speed and quality of public services, particularly in social welfare systems (Berryhill et al., 2019). Through the adoption of various AI solutions, governments aim to leverage the specific advantages this technology offers for welfare services (Figure 4). One of the most significant outcomes of AI implementation is the marked improvement in public service delivery. This is especially evident in the increased responsiveness, efficiency and cost-effectiveness of welfare services, a benefit widely recognised across the public sector (44 cases). AI's role in streamlining operations and delivering personalised support underscores a strong focus on enhancing the immediate experience of those who rely on social welfare, with public (citizen)-centred services receiving significant attention (28 cases), alongside the increased quality of public sector information and services (27 cases) and personalised services (25 cases) with the introduction of new services or channels also noted (22 cases). Moreover, a substantial emphasis is placed on improving administrative efficiency, with numerous mentions highlighting the enhanced quality of processes and systems (33 cases) and the greater responsiveness of government operations (23 cases). Improved management of public resources is also a key area of focus (18 cases), while the remaining areas of potential received less emphasis. These points underline AI's crucial role in optimising internal governmental operations, assuring that social welfare services are not only improved but also delivered more efficiently and effectively. However, there is a noticeable gap in the attention given to open government capabilities, such as public participation and transparency. This aspect is less frequently emphasised, with only 10 AI use cases focused on increasing transparency in public sector operations and just 1 AI use case aimed at improving public control and influence over government actions and policies. Moreover, there is no AI use case dedicated to enhancing public participation in government actions and policymaking. This suggests that while AI is being used to improve the delivery and internal efficiencies of welfare services, its potential to empower citizens and strengthen democratic processes within the context of social welfare remains underdeveloped.

Figure 4: Number of AI use cases in social protection across the EU member states

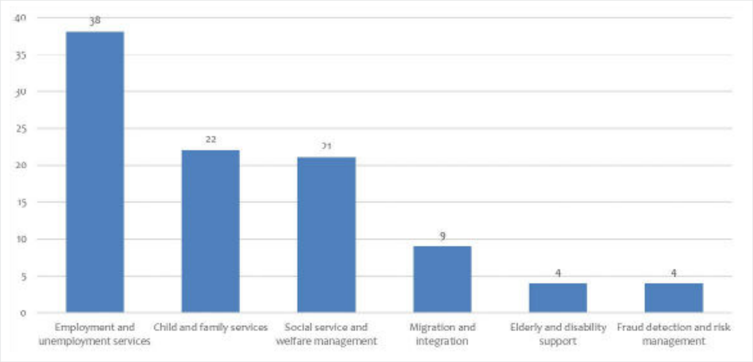


Source: European Commission, 2024

The most prominent use of AI technology has been in employment and unemployment services.

The adoption of AI within the social protection framework across the EU public sector showcases the clear prioritisation of certain areas over others. Specifically, the most prominent use of AI technology has been in employment and unemployment services. With 38 documented cases, this area stands out as the primary focus, likely due to the significant role employment plays in economic stability and individual well-being. AI's integration here could involve automated job matching, predictive analytics for job market trends, or boosting the efficiency of unemployment benefits processing, reflecting the sector's need for innovative solutions to address complex labour market dynamics. In addition to employment services, AI has made substantial inroads in child and family services as well as social service and welfare management, with 22 and 21 use cases, respectively. These areas are critical for maintaining social welfare, particularly for vulnerable populations. AI applications here might include case management systems that use predictive analytics to identify at-risk children or families, tools that optimise the allocation of social benefits, or platforms that streamline the coordination between different social services. The significant number of use cases suggests that AI is increasingly seen as a tool to add to the effectiveness and efficiency of social welfare systems, potentially leading to better outcomes for those in need. Migration and integration services have seen a moderate level of AI adoption, with 9 documented use cases. This could involve the use of AI to assist in processing asylum applications, managing migration flows, or providing personalised support to migrants as they integrate into new communities. The moderate level of AI use in this area may reflect the complexities and sensitivities involved in migration issues, where AI can assist, but human oversight remains crucial. On the other hand, elderly and disability support, along with fraud detection and risk management, have seen relatively limited AI adoption, each with only 4 cases. The lower implementation in elderly care and disability support could be due to the challenges of applying AI in contexts that call for high levels of personalised care and human interaction. In these areas, AI might be used for specific tasks, such as monitoring health conditions or assisting with daily activities, but its role remains supplementary rather than central. Fraud detection and risk management also have a smaller number of AI implementations, possibly because these areas require highly specialised AI tools capable of analysing vast amounts of data to identify potential fraud or risks. The low number of use cases might indicate that while AI holds considerable potential in these areas, it is still in the early stages of development or deployment.

Figure 5: Area of AI use cases in social protection across the EU public sector



Source: European Commission, 2024

Overall, the findings illustrate that AI's role in social protection across the EU public sector is largely focused on improving efficiency and outcomes in employment and welfare-related services. However, its application in areas such as elderly care, disability support, and fraud detection is still emerging, suggesting untapped potential or existing barriers to broader AI integration in these sectors. Further details and examples of key areas where AI is being utilised within social protection across the EU public sector are presented in Table 2.

Table 2: Key areas of using AI in social protection across the EU public sector

Area	Description	Examples
Employment and Unemployment Services	Initiatives designed to assist job seekers, manage unemployment, and match individuals with job opportunities using data and AI	Job matching platforms, unemployment benefit systems, workforce training programmes
Child and Family Services	Efforts focused on supporting children and families through predictive modelling, risk assessment, and proactive social service delivery	Early intervention programmes, child welfare monitoring, family support services
Social Service and Welfare Management	Activities focused on improving the management and delivery of social services and welfare through automated systems and data-driven approaches	Welfare programme administration, social service delivery platforms, benefit eligibility assessments
Migration and Integration	Actions addressing immigration and migrant integration, utilising digital tools and AI to streamline processes and support social services	Immigration case management, integration support services, migrant assistance programmes

Area	Description	Examples
Elderly and Disability Support	Activities aimed at enhancing services for the elderly and individuals with disabilities, ensuring personalised care and improved quality of life	Assistive technology for the elderly, disability benefits management, elderly care coordination
Fraud Detection and Risk Management	Efforts to identify and mitigate fraud in social services and welfare systems, using advanced analytics and machine learning	Fraud detection in welfare payments, risk assessment for social benefits, compliance monitoring

Source: European Commission, 2024

6 Conclusion

The integration of AI into welfare systems is a pivotal transformation in the way public services are delivered, as vividly demonstrated by the Slovenian case and further supported by a comprehensive examination across the European Union. This shift is not merely a technological advancement; it symbolises a reimagining of how societies can better serve their most vulnerable populations. AI holds the potential to revolutionise welfare systems by making them more efficient, transparent and equitable. It can streamline administrative processes, reduce human error, and ensure that resources are allocated more fairly. In any case, these benefits are not automatic and must be carefully cultivated. The successful integration of AI into welfare systems requires a thoughtful, strategic approach that addresses the inherent challenges of deploying AI in such sensitive domains. Policymakers hold the key to navigating these challenges as they are tasked with creating and enforcing policies that assure AI serves the public good while rigorously safeguarding individual rights. From this careful consideration of the Slovenian example and broader EU practices, several critical recommendations for policymakers emerge, each aimed at maximising the benefits of AI while minimising its risks.

> A fundamental step for policymakers is to develop and enforce comprehensive legal and regulatory frameworks that govern AI use in welfare systems.

Strengthening Legal and Regulatory Frameworks: a fundamental step for policymakers is to develop and enforce comprehensive legal and regulatory frameworks that govern AI use in welfare systems. These frameworks must not only comply with existing laws, such as the EU AI Act, but also be tailored to the specific needs of welfare services. Clear guidelines are needed for the deployment and operation of AI systems to make sure they promote social equity and respect fundamental rights. The Slovenian case underscores the importance of such frameworks in preventing errors and ensuring that AI systems adhere to national legal standards. These frameworks should also include mechanisms for regular updates and adaptations, reflecting the evolving nature of AI technology and its impact on welfare.

Incorporating Human Oversight: while AI can automate many administrative processes, the human element remains indispensable in welfare services. Although AI excels at processing large volumes of data and making rapid decisions, it lacks the empathy and contextual understanding that human oversight provides. The Slovenian experience illustrates that while AI can enhance efficiency, it cannot replace the nuanced judgment required in welfare services. Policymakers should ensure that welfare systems incorporate human intervention points, especially in complex or sensitive cases. This approach helps AI support, rather than replace, the compassionate and personalised service essential in welfare systems. Moreover, human oversight is crucial for identifying and correcting errors or biases that may arise from AI-driven processes.

Ensuring Transparency and Accountability: transparency is key to maintaining public trust, particularly in critical systems like welfare. AI systems can be complex and opaque, making it difficult for individuals affected by AI decisions to understand how these decisions are made. The Slovenian case highlights the need for transparent AI processes, as seen in the challenges faced during the implementation of e-Welfare. Policymakers must make sure that AI systems used in welfare are not only explainable but also designed with transparency in mind. Beneficiaries should have clear insights into how decisions about their welfare are made, with AI processes being fully auditable. In addition, there must be well-defined accountability structures, ensuring clear channels for redress and accountability when things go wrong.

Prioritising Data Protection and Privacy: welfare systems handle some of the most sensitive personal data, including financial, health and social information. The integration of AI adds to the complexity of data processing, raising significant privacy concerns. As demonstrated by the Slovenian case, any lapses in data management can lead to significant public distrust and operational failures. Policymakers must prioritise stringent data protection measures to ensure AI systems comply with regulations like the GDPR. This includes implementing robust security protocols to protect data from breaches and unauthorised access. Beyond compliance, welfare policies should advocate for data minimisation practices, assuring that only necessary data is collected and processed. Protecting individuals' privacy is not just about legal compliance; it is essential for maintaining the trust of citizens who rely on these systems.

Mitigating Bias and Ensuring Fairness: AI systems, despite their power, are prone to biases that can lead to unfair outcomes, especially for marginalised groups. Experiences across the EU, including challenges faced in Slovenia, demonstrate the risks of unchecked AI deployment. Policymakers must focus on creating policies that require regular audits and the continuous monitoring of AI systems to detect and mitigate biases. This includes implementing fairness checks at multiple stages of the AI lifecycle and providing clear pathways for individuals to challenge decisions they believe are unjust. Ensuring fairness in AI-driven welfare systems is essential for upholding the principles of social justice and equality.

In conclusion, the integration of AI into welfare systems, as evidenced by the Slovenian case and the broader EU experience, offers an unprecedented opportunity to transform public service delivery in ways previously unimaginable. However, the full realisation of AI's potential in welfare systems is contingent upon the careful, deliberate actions of policymakers who must prioritise a range of critical factors. These include the establishment of robust legal and regulatory frameworks that are both comprehensive and adaptable, the incorporation of essential human oversight to complement AI's capabilities, the enforcement of transparency and accountability measures to maintain public trust, and the stringent protection of data privacy to safeguard sensitive personal information. Further, making sure that AI systems operate with fairness and are free from biases is crucial for upholding the principles of social justice and equality. By focusing on these key areas, policymakers can ensure that AI not only enhances the efficiency of welfare systems but also reinforces the core values of equity, compassion and trust that are fundamental to social welfare. This approach will help to build a future in which AI serves as a powerful tool in the promotion of social good, creating welfare systems that are not only smarter but also more humane.

REFERENCES

- Algorithm Watch (2020b). In Spain, the VioGén algorithm attempts to forecast gender violence. Retrieved 31 July 2024 from: https://algorithmwatch.org/en/viogen-algorithm-gender-violence/
- Algorithm Watch. (2020a). How Dutch activists got an invasive fraud detection algorithm banned. Retrieved 31 July 2024 from: https://algorithmwatch.org/en/syri-netherlands-algorithm/
- Amnesty International. (2021). Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal Retrieved 31 July 2024 from: https://www.amnesty.org/en/documents/eur35/4686/2021/en/
- Ananny, M. (2016). Toward an ethics of algorithms: Convening, observation, probability, and timeliness. Science, Technology, & Human Values, 41(1), 93-117. https://doi.org/10.1177/0162243915606523
- Auby, J.-B. (Ed.) (2014). Codification of Administrative Procedures. Bruylant.
- Babšek, M., & Kovač, P. (2023a). Social procedures and artificial intelligence: opportunities and barriers in Slovenia and beyond. In: The future of public administration enabled through emerging technologies: 31th NISPAcee Annual Conference, May 25-27, 2023, Belgrade, Serbia: e-proceedings. Retrieved 4 July 2024 from: https://www.nispa.org/files/conferences/2023/e-proceedings/system_files/papers/NISPAcee23_BabsekKovac_apr23.pdf
- Babšek, M., & Kovač, P. (2023b). The Covid-19 Pandemic as a Driver of More Responsive Social Procedures: between Theory and Practices in Slovenia. NISPAcee Journal of Public Administration and Policy, 16(1), 1-32. https://doi.org/10.2478/nispa-2023-0001
- Berryhill, J., Heang, K. K., Clogher, R., & McBride, K. (2019). Hello, World: Artificial intelligence and its use in the public sector. OECD Working Papers on Public Governance. No. 36. OECD Publishing: Paris. https://doi.org/10.1787/726fd39d-en
- Coglianese, C. (2021). Administrative law in the automated state. Daedalus, 150(3), 104-120. http://dx.doi.org/10.1162/daed_a_01862
- Digital Watch. (2019). Discriminatory employment algorithm towards women and disabled. Discriminatory employment algorithm towards women and disabled. https://dig.watch/updates/discriminatory-employment-algorithm-towards-women-and-disabled
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., ... & Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 57, 101994. https://doi.org/10.1016/j.ijinfomgt.2019.08.002
- Edwards, L. (2021). The EU AI Act: a summary of its significance and scope. Artificial Intelligence (the EU AI Act), 1. Retrieved 2 July 2024 from: https://www.adalovelaceinstitute.org/wp-content/uploads/2022/04/Expert-explainer-The-EU-AI-Act-11-April-2022.pdf
- Eticas. (2022). The External Audit of the VioGén System. Retrieved 31 July 2024 from: https://eticasfoundation.org/wp-content/uploads/2024/07/ETICAS-FND-The-External-Audit-of-the-VioGen-System-1-1.pdf

- European Commission. (2021). Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial intelligence Act) and amending certain Union legislative acts, COM(2021) 206 final, 2021/0106(COD). Brussels. Retrieved 2 July 2024 from: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52021PC0206
- European Commission. (2024). Public Sector Tech Watch latest dataset of selected cases. European Commission [Dataset]. Retrieved 4 July 2024 from: http://data.europa.eu/89h/e8e7bddd-8510-4936-9fa6-7e1b399cbd92
- European Commission. (2024). Public Sector Tech Watch latest dataset of selected cases. European Commission [Dataset]. PID: http://data.europa.eu/89h/e8e7bddd-8510-4936-9fa6-7e1b399cbd92
- European Parliament, & European Council. (2024). Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), 2021/0106(COD). Retrieved 4 July 2024 from: https://data.consilium.europa.eu/doc/document/PE-24-2024-INIT/en/pdf
- European Parliament. (2024). EU AI Act: first regulation on artificial intelligence. Retrieved 4 July from: https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#ai-act-different-rules-for-different-risk-levels-0
- Haldar, M., & Larsson, K. K. (2021). Can computers automate welfare? Norwegian efforts to make welfare policy more effective. Journal of Extreme Anthropology, 5(1), 56-77. https://doi.org/10.5617/jea.8231
- Hassan, B. (2022). Artificial intelligence in social security: Opportunities and challenges. Журнал исследований социальной политики, 20(3), 407-418. https://doi.org/10.17323/727-0634-2022-20-3-407-418
- Henman, P. (2020). Improving public services using artificial intelligence: possibilities, pitfalls, governance. Asia Pacific Journal of Public Administration, 42(4), 209-221. https://doi.org/10.1080/23276665.2020.1816188
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who's the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. Business horizons, 62(1), 15-25. https://doi.org/10.1016/j.bushor.2018.08.004
- Kaun, A. (2022). Suing the algorithm: the mundanization of automated decision-making in public services through litigation. Information, Communication & Society, 25(14), 2046-2062. https://doi.org/10.1080/1369118X.2021.1924827
- Kuhlmann, S., & Wollmann, H. (2019). Introduction to Comparative Public Administration: Administrative Systems and Reforms in Europe. Edward Elgar Publishing.
- Misuraca, G., & Van Noordt, C. (2020). AI Watch-Artificial Intelligence in public services: Overview of the use and impact of AI in public services in the EU. JRC Research Reports, (JRC120399). https://dx.doi.org/10.2760/039619
- Misuraca, G., Barcevicius, E., & Codagnone, C. (2020). Exploring Digital Government Transformation in the EU–Understanding public sector innovation in a data-driven society (JRC121548). Joint Research Centre. https://dx.doi.org/10.2760/480377
- Ohlenburg, T. (2020). AI in Social Protection – Exploring Opportunities and Mitigating Risks. Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ) GmbH. Retrieved 31 July 2024 from: https://socialprotection.org/discover/publications/ai-social-protection
- Panigutti, C., Hamon, R., Hupont, I., Fernandez Llorca, D., Fano Yela, D., Junklewitz, H., ... & Gomez, E. (2023). The role of explainable AI in the context of the AI Act. In Proceedings of the 2023 ACM conference on fairness, accountability, and transparency (pp. 1139-1150).

- Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2020). Big Data and AI–A transformational shift for government: So, what next for research? Public Policy and Administration, 35(1), 24-44. https://doi.org/10.1177/0952076718780537
- Politico. (2022). Dutch scandal serves as a warning for Europe over risks of using algorithms. Retrieved 31 July 2024 from: https://www.politico.eu/article/dutch-scandal-serves-as-a-warning-for-europe-over-risks-of-using-algorithms/
- Ranchordas, S. (2022). Empathy in the digital administrative state. Duke Law Journal, 71, 1341-1389. Retrieved 4 July 2024 from: https://scholarship.law.duke.edu/dlj/vol71/iss6/4
- Sekwenz, M.-T. (2022). Algorithmic Bias in the Public Sector – A View from Austria. Retrieved 31 July 2024 from: https://www.ippi.org.il/algorithmic-bias-in-the-public-sector-a-view-from-austria/
- Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 36(2), 368-383. https://doi.org/10.1016/j.giq.2018.09.008
- Telefónica. (2024). A fit for purpose and borderless European Artificial Intelligence Regulation. Retrieved 2 July 2024 from: https://www.telefonica.com/en/communication-room/blog/a-fit-for-purpose-and-borderless-european-artificial-intelligence-regulation/
- Van Bekkum, M., & Borgesius, F. Z. (2021). Digital welfare fraud detection and the Dutch SyRI judgment. European Journal of Social Security, 23(4), 323-340. https://doi.org/10.1177/13882627211031257
- Wang, C., Teo, T. S., & Janssen, M. (2021). Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector. International Journal of Information Management, 61, 102401. http://dx.doi.org/10.1016/j.ijinfomgt.2021.102401
- World Economic Forum. (2017). 2017 is the year of artificial intelligence. Here's why. Retrieved 31 July 2024 from: https://www.weforum.org/agenda/2017/05/2017-is-the-year-of-artificial-intelligence-here-s-why/

Chapter 11

Globally Applicable Gold Standard: Exploring the Extraterritorial Implications of the AI Act on International Arbitration

Jasmina Ibrahimpašić

1 Introduction

The General Data Protection Regulation (GDPR) came into effect on 25 May 2018 and today represents the strongest privacy and cybersecurity law in the world. What makes the GDPR the toughest law in the world is the extraterritorial scope it derives from Article 3, which imposes obligations on anyone anywhere so long as their actions are aimed at the collection and processing of private data from natural persons in the European Union (EU). Six years later, the EU is doing it again – but with the regulation of Artificial Intelligence (AI).

Adopted by the European Parliament on 13 March 2024 and entered into force on 1 August 2024, the AI Act is the first globally comprehensive legal framework on AI in the world. The act aims to provide clear requirements and obligations concerning AI use. What differentiates the AI Act from the GDPR is the European AI Office – a key player in implementation of the AI Act, the promotion of trustworthy AI, and ensuring a strategic, coherent and effective European approach to AI on the international level. However, the most important role is to position Europe as the global leader in the regulation of AI.

Drawing parallels with the GDPR's global reach, it is certain that the international arbitration community will have to adhere to the AI Act. Still, to what extent the AI Act will force a change in procedural rules and regulatory frameworks within the international arbitration community has yet to be assessed along with whether such a change will ensure harmonised AI governance in international arbitration. In this regard, since many arbitration rules already incorporate GDPR provisions and principles to protect private and personal data, arbitrators and parties involved may need to ensure compliance with both the GDPR and the AI Act. This could lead to disclosure requirements regarding the use of AI tools in all phases of arbitral proceedings, with a risk of an arbitral award not being compliant with the principles upheld by the United Nations Convention on the Recognition and Enforcement of Foreign Arbitral Awards (The New York Convention), in turn making the arbitral award unenforceable. As international arbitration is constantly developing and evolving with the progression into the world of Big Data and AI, it is a fascinating time to observe which changes will stir things up.

2 European Approach to Artificial Intelligence

On 25 April 2018, the European Commission issued the document Artificial Intelligence for Europe, also known as the European AI Strategy. In this document, the European Commission defined AI as "systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals" (EUR-Lex - 52018DC0237 - EN - EUR-Lex, 2018). The European Commission further emphasised the need for a coordinated approach to the use of AI and all the new challenges such use would bring. In this sense, the European Commission especially focused on ethical and legal questions regarding liability or potentially biased decision-making. Such issues were also voiced within the international arbitration community due to the growing reliance on AI tools in the arbitral process.

The European Commission laid out the aims of a coordinated plan for AI. In light of this, the document hints at three main aims which will become more defined in subsequent documents from the European Commission. Yet, in this document the European Commission expressed its will to maximise the impact of investments, on both European and national levels, to encourage synergies and cooperation across the EU and collectively define the way forward in which the EU can compete globally as a whole (EUR-Lex - 52018DC0237 - EN - EUR-Lex, 2018).

Then, on 7 December 2018 the European Commission published the Coordination Plan on Artificial Intelligence aimed at the European Parliament, the European Council, the European Economic and Social Committee, and the Committee of the Regions (EUR-Lex - 52018DC0795 - EN - EUR-Lex, 2018). This document presents an effort to maximise Europe's potential to compete globally in the uptake and development of AI in different sectors and is a joint commitment between the European Commission, member states, Norway and Switzerland (Coordinated Plan on Artificial Intelligence | Shaping Europe's Digital Future, 2023).

The Coordinated Plan on AI provided a strategic framework for the further development of national AI strategies (EUR-Lex - 52018DC0795 - EN - EUR-Lex, 2018). With this, the member states were encouraged to dive into the process of drafting their national strategies to regulate the uptake, development and use of AI across different sectors. After a year and a half, the European Commission presented its AI package which included its Communication on fostering a European Approach to AI, the 2021 review of the Coordinated Plan on AI, and its Regulatory framework proposal on artificial intelligence (European Commission, 2023).

On 21 April 2021, the Coordinated Plan on AI was revised and updated. Now known as the 2021 Coordinated Plan on AI, it was an extension of the collaboration from 2018 between the European Commission and member states (European Commission I, 2021). The 2021 review of the Coordinated Plan on AI focused on the EU's ability to accelerate, act and align its AI policy priorities and investments. In this context, the 2021 Coordinated Plan proposed three key actions needed to jumpstart the EU as a leader of trustworthy AI. Those were the need to: i) accelerate investment in AI technologies; ii) act on AI strategies and programmes fully and in a timely manner; and iii) align AI policy to eliminate fragmentation and address global challenges.

Alongside the 2021 Coordinated Plan, the European Commission announced the Proposal for a Regulation on AI, also known as the AI Regulation, which laid down harmonised rules on AI. The AI Regulation addressed the risks of specific uses of AI and categorised them on four different levels – unacceptable risk, high risk, limited risk and minimal risk (European Commission II, 2021). These documents served as the foundation for the development of the AI Act.

3 Different Regulatory Approaches to Artificial Intelligence in Austria, France, and Germany

It is important to note that, thanks to the Coordinated Plan on AI from 2018, many member states came forward with their own national AI strategies. To date, all 27 member states have their national AI strategy with Croatia being one of the last countries whose National AI Strategy was published in 2022

(Republic of Croatia, 2022). For the purposes of this article, we focus on three member states – Austria, France and Germany – since within these countries lie some of the most famous arbitral institutions.

3.1 Austria's Regulatory Approach to Artificial Intelligence

According to OECD.AI data, Austria currently has a total of 9 policies regarding AI (OECD AI Policy Observatory Portal III, 2024). In 2017, Austria introduced a Digital Roadmap with 117 measures and 36 digitalisation principles to redesign digitalisation in Austria (Digital Austria Act – Das Digitale Arbeitsprogramm Der Bundesregierung, 2023). In August of the same year, the Austrian Council for Robotics and AI (ACRAI) was established by the Ministry for Transport, Innovation and Technology (Austria AI Strategy Report - European Commission, 2021; OECD AI Policy Observatory Portal II, 2024). ACRAI consisted of experts on robotics and AI from the research field, academia and industry with the task to help the Ministry develop an AI strategy. It had two main objectives – to provide an Austrian Strategy for Robotics and AI on one hand and, on the other, to give general advice concerning ethical, economic, technological and legal aspects of AI. However, on 1 July 2023 ACRAI was dissolved according to the underlying FWIT Council Establishment Act and transferred to the newly established Council for Research, Science, Innovation and Technology Development (FORWIT, 2023).

In 2019, the Federal Ministry for Transport, Innovation and Technology and the Federal Ministry for Digital and Economic Affairs of Austria introduced the AI Mission Austria 2030 (AIM AT 2030). This federal AI strategy is aimed at setting the framework for the responsible use of AI in various fields such as research and innovation, qualification and training for citizens regarding AI use, use of AI in the public sector, the economy and labour market. The central objectives of the AIM AT 2030 Strategy are to promote the responsible and broad use of AI in the public interest based on European fundamental values, develop measures to recognise, mitigate and prevent possible dangers resulting from AI, raise the research, technology and innovation of AI, position Austria as a leading research and innovation location for AI worldwide, and create a legal framework that ensures safety in the use of AI that meets the requirements of the EU legislation (OECD AI Policy Observatory Portal I, 2023). Apart from these, other AI policies in Austria include the Automated Driving Regulation, the testing of automated driving on public roads, and the Platform Industry 4.0 which provides a hub supporting policy coordination between relevant stakeholders.

In the international arbitration community, Austria is known for the Vienna International Arbitral Centre (VIAC), which is one of Europe's leading arbitral institutions, founded in 1975 as a department of the Austrian Federal Economic

Chamber. Currently, VIAC is a premier international arbitral institution for the settlement of commercial disputes in both the regional and international community (About Us – Vienna International Arbitral Centre, 2024).

3.2 France's Regulatory Approach to Artificial Intelligence

As reported by OECD.AI data, France has 35 AI policies (AI Strategies and Policies in France, 2024). Even though the AI Act is expected to fulfil its function, France is likely to "regulate AI on a sector-specific basis" (AI Watch: Global Regulatory Tracker – France | White & Case LLP, 2024).

In March 2018, the French politician and mathematician Cédric Villani released a report called "For a Meaningful Artificial Intelligence: Towards a French and European Strategy", a task assigned to him by Prime Minister Édouard Philippe. His research focused on building a data-focused economic policy, promoting agile and enabling research on mathematics and AI, assessing the effects of AI on the future of work and the labour market, the use of AI for a more ecological economy, the ethical challenges and considerations of AI, and making AI inclusive and diverse (Villani, 2018).

Following this report and the framework laid out in the France Plan 2030, the French government launched the national strategy for AI called SNIA in 2018 with a view to "position France as one of the European and world leaders in the field of artificial intelligence" (La Stratégie Nationale Pour l'IA | Entreprises.gouv.fr, 2023) The SNIA is divided into two phases. Phase one was from 2018 to 2022 and targeted equipping France with competitive research capabilities, while phase two is from 2021 to 2025 and seeks to disseminate technologies of AI across the economy (AI Watch: Global Regulatory Tracker – France | White & Case LLP, 2024; La Stratégie Nationale Pour l'IA | Entreprises.gouv.fr, 2023).

This led to several calls for AI projects within the country. Such projects were the Call for Trusted Artificial Intelligence Demonstrators (DIAC), the second wave of the Demonstrators of Artificial Intelligence in Territories (DIAT) call for projects, the Relaunch of the Technological Maturation and Demonstration of Embedded Artificial Intelligence Technologies call for projects, the call for Digital Commons for Generative AI projects, the call for AI Booster France 2030 projects and the AI cluster call for projects. However, it is worth noting that on 19 September 2023 the Generative AI Committee was inaugurated to assemble all relevant stakeholders from cultural, economic and technological sectors to help the French government in making France a leader of the AI revolution (AI Watch: Global Regulatory Tracker – France | White & Case LLP, 2024).

In the international arbitration community, France is known for the International Chamber of Commerce (ICC), founded in 1919 in the aftermath of World War

One (Our Mission, History and Values, 2024). The founders of the ICC were a group of industrialists, financiers and traders who called themselves "Merchants of Peace" and believed the private sector is best qualified to set global standards for business. Today, the ICC has many national committees around the world and works to secure peace, prosperity and opportunity for everyone.

3.3 Germany's Regulatory Approach to Artificial Intelligence

According to OECD.AI data, Germany has 42 policies (AI Strategies and Policies in Germany, 2024). On 15 November 2018, the German Federal Government announced the National Strategy for Artificial Intelligence, which was jointly developed by the Federal Ministry of Education and Research, the Federal Ministry for Economic Affairs and Energy and the Federal Ministry of Labour and Social Affairs (OECD AI Policy Observatory Portal IV, 2024; AI Strategies – Home, 2023; Germany AI Strategy Report, 2021). The strategy observed the progress made in the use and development of AI, potential goals to be achieved in the future and provided a concrete plan of policy actions. The main goals were to make Germany and Europe a leading centre for AI by attracting investments, guarantee the responsible development of AI, and integrate AI into society in ethical, legal, cultural and institutional terms.

In 2019, the German Federal Government published an interim report on the implementation of measures of the Strategy for AI and presented its ethical guidelines and specific recommendations on AI concerning data usage and algorithm-based decision-making. The Strategy for AI was reviewed in December 2020 "to strengthen Germany as an internationally competitive centre of AI research, development and application" (The German Federal Government, 2020).

Although Germany does not have any particular laws that directly regulate AI, as is expected of the AI Act, in mid-2021 Germany included a reference to AI in three provisions of its Works Constitution Act, specifically in Section 80(3), Section 90(1) No. 3 and Section 95(2a) (AI Watch: Global Regulatory Tracker – Germany | White & Case LLP, 2024). All three provisions are related to the right to information for work councils when AI is to be used in the workplace. In Section 80(3) of the Works Constitution Act, the focus is on the introduction or application of AI when there is a need for it to carry out the tasks of the works council, in which case the calling in of an expert is necessary. In Section 90(1) No. 3, focus is given to the employer and its obligation to inform the work council about the use of AI in work processes and workflows, while in Section 95(2a) the focus is on the approval of the works council when AI was used in the drawing up of the guidelines for professional and personal requirements in the workplace (Works Constitution Act (Betriebsverfassungsgesetz – BetrVG), 2022).

In the international arbitration community, Germany is known for its German Arbitration Institute (DIS) whose roots can be traced back to 1920 when the German Arbitration Committee (DAS) was established in Berlin (German Arbitration Institute (DIS): Our History, 2024). Since being founded, DIS has been the leading arbitral institution in Germany for the administration of arbitrations and other alternative dispute resolution proceedings for national and international commercial disputes.

4 The International Arbitration Community's Approach to Artificial Intelligence

Over the course of recent decades there have been many technological developments comparable to those of the Industrial Revolution. However, the rapid development of AI in this period raises questions regarding legal, ethical and philosophical challenges.

As far as the legal field is concerned, the use of AI within the legal community is becoming a common occurrence in all areas of law. The use of AI tools has already seriously affected how legal business is conducted, how contracts and transactions are entered into, and disputes are raised and resolved (Berardicurti, 2021). For example, a survey in France that included lawyers, notaries, legal officers and experts "found that 80 per cent of legal professionals believe that generative AI tools will enhance their work efficiency" (Juhan, et al., 2024, p. 15). This is also applicable to the international arbitration community because AI tools have become common in almost all phases of the arbitral process.

Today, AI tools are used for various tasks within the legal industry, from practice management, conflict management, contract review and due diligence to legal assistance, e-discovery reviews and outcome prediction. The different use and applicability in various stages of the legal process means AI tools can be placed in four categories: "1) simple AI tools used for accurate and efficient legal research; 2) AI tools used for selecting suitable experts, counsels and arbitrators; 3) AI tools used for facilitating procedural automation by translating, transcribing, summarising evidence and even drafting compilatory parts of legal documents and arbitral awards; and 4) AI tools used for their use in the adjudication process (including the "tools of predictive justice") (Chauhan, 2020). Most of the AI tools described above have proven themselves useful for reducing the time and costs of arbitration and therefore correlate to arbitration's greater efficiency.

In the context of the above-mentioned categorisation of AI tools, selection and predictive AI tools raise the most concern within the arbitration community given their potential impact on the arbitrator's role and function, due process concerns and the fundamental rights of the parties (Carrara, 2020). This means we must consider the risk-based approach of the AI Act, which defines four levels of risk for AI systems, unacceptable risk, high risk, limited risk and minimal risk (European Commission, 2024), when discussing the use of AI tools in international arbitration. In this context, it is important to differentiate automated from autonomous systems. According to one definition, "autonomous systems are those which can take decisions themselves without being explicitly programmed, whereas automated systems must follow a pre-determined set of instructions with no discretion as to how they are to be followed" (Waqar, 2022, pp. 346–347).

With regard to the potential use of AI tools in arbitral proceedings, three scenarios can be considered. In the first scenario, which refers to the present day, the use of technology in arbitral proceedings is still in its primal stages. The use of AI in international arbitration is dependent on and limited by the quality of the data processed and the algorithm applied. The second scenario explores the possibility of increased usage of AI within the international arbitration community. Here, the use of AI would be similar to the first scenario since AI would be used as a strictly instructional method rather than a decision-making method. In this scenario, AI would be used to read a contract and identify the existence of an arbitration clause, further distinguishing lex arbitri from lex loci arbitri, to translate hundreds of documents for greater efficiency of the process and to reduce time and, finally, to draft an abstract of the arbitration award (Waqar, 2022). These two scenarios demonstrate the use of AI tools as automated systems.

The third scenario explores the possibility of introducing the Robot Arbitrator (Scherer, 2022), an AI system able to make its own decisions through its own cognitive and analytical process. A milder version of this is Kira Systems, machine learning software used by many London firms to analyse legal documents (Waqar, 2022). However, a more powerful version of this would be the experiment being jointly conducted by CIArb Brazil Branch, Willem C. Vis International Commercial Arbitration Moot (Vis Moot) and VIAC on how AI can perform as a decision-maker (CIArb – Brazil, 2024).

The practical experiment included finalists from the 30th Vis Moot Competition from the University of Belgrade and the University of Vienna and a tribunal of three arbitrators, where two of the three were ChatGPT 4.0 bots. Results of the experiment showed that AI can reach the same result as an arbitrator if fed with the correct data, and further demonstrated the "enormous potential to become an indispensable tool in international arbitration" (CIArb – Brazil, 2024).

Although the AI Act has become the new reality, the international arbitration community remains quiet on this topic. Nonetheless, considering the extraterritorial effect of Article 3 of the GDPR, by analogy Article 2 of the AI Act is likely to push arbitral institutions to revise their rules with respect to the use of AI tools in the arbitration process.

5 Potential Applicability of the AI Act to International Arbitration

When the GDPR came into force on 25 May 2018, the first question the international arbitration community had was whether the GDPR applies to international arbitral proceedings (Gordon, 2020). Namely, the GDPR's scope of applications was assessed to determine in what manner the GDPR would apply to international arbitration. The GDPR's scope of application is known as the personal, material and territorial scope of the GDPR and "defines whose personal data is covered, what personal data is protected, and considers the location of the data subject and data controller during processing activities" (The Ultimate Guide to the EU General Data Protection Regulation (GDPR), 2022).

Article 2 of the GDPR provides the material scope of the regulation, which is the processing of personal data, while Article 3 explains the regulation's territorial scope. Article 3 stipulates that the GDPR applies to the processing of personal data that happens within the establishment of a controller or processor in the EU, regardless of whether the processing occurs in the EU or somewhere else. In this context, the GDPR demonstrated the power of EU regulations to influence global markets. "Prompted by the entry into force of the GDPR, consumers were flooded with e-mails from numerous companies, asking for their consent on the use of their personal data" (Door Anu Bradford, 2020, p. 169).

The GDPR also laid out relevant definitions, including of controller and processor. In Article 4(7), the GDPR defines a controller as a natural or legal person who alone or jointly with others determines the purposes and means of the processing of personal data, while paragraph (8) of the same Article defines a processor as a natural or legal person which processes personal data on behalf of the controller. Based on this, Gordon (2020) explored the GDPR's applicability to arbitrators and expressed the possibility of arbitrators subjected to the GDPR being qualified as a controller pursuant to Article 4(7) of the GDPR.

Therefore, to assess whether the AI Act applies to international arbitration it is important to look at applicability with respect to its material, personal and territorial scope. In the context of material and territorial scope, Article 2 of the

AI Act plays a significant role. In relation to its material scope, the AI Act applies when AI systems are placed on the market or put into service in the EU. As a result, some activities of arbitrators who rely on AI tools could be classified as high-risk. This is because in Annex III, which deals with high-risk AI systems from Article 6(2) of the AI Act, Article 8(a) states that AI systems "intended to be used by a judicial authority or on their behalf to assist a judicial authority in **researching and interpreting facts and the law** and in **applying the law to a concrete set of facts,** or to be **used in a similar way in alternative dispute resolution** (emphasis added)" are high-risk (Scherer, 2024).

This is confirmed by Recital 61, which states that AI systems intended for the administration of justice and democratic processes should be classified as high-risk (*(61) - EU AI Act,* 2024). The reasoning behind this is the considerable potential of these AI systems to significantly impact the democracy, the rule of law, the right to an effective remedy and to a fair trial and, therefore the need to address the challenges of potential biases, errors and opacity. Recital 61 also refers to alternative dispute resolution and equates it with judicial authority because the outcomes of such proceedings produce legal effects for the parties involved.

It is also evident from Recital 61 that regulators had international arbitration in mind since they also refer to AI systems intended to be used by alternative dispute resolution bodies when stating that such AI systems "should also be considered to be high-risk when the outcomes of the alternative dispute resolution proceedings produce legal effects for the parties" (*(61) - EU AI Act,* 2024). However, this classification will not affect AI systems used for ancillary administrative activities that do not influence the actual administration of justice in particular cases.

Despite the high-risk classification, Article 6(3) of the AI Act contains exceptions for AI systems that do not pose a risk of harm to the health, safety or fundamental rights of natural persons, including not materially influencing the outcome of decision-making. This applies to the following situations:

a. The AI system is intended to perform a narrow procedural task;

b. The AI system is intended to improve the result of a previously completed human activity;

c. The AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns and is not meant to replace or influence the previously completed human assessment, without proper human review; or

d. The AI system is intended to perform a preparatory task to an assessment relevant for the purposes of the use cases listed in Annex III".

Scherer (2024) states that it is unclear from the start in which circumstances these expectations apply and whether international arbitration will fall into the high-risk activities category only when natural persons are concerned. However, Recital 53 provides an answer to this question. The recital affirms that some specific cases may emerge in which AI systems do not lead to significant harm to legal interests because they do not materially influence the decision-making. In this sense, "an AI system that does not materially influence the outcome of decision-making should be understood to be an AI system that does not have an impact on the substance, and thereby the outcome, of decision-making, **whether human or automated"** (emphasis added) (*(53) - EU AI Act,* 2024).

This recital further provides several conditions when the AI system does not materially influence the outcome of decision-making: 1) the AI system is intended to perform a narrow procedural task (classifying incoming documents into categories or detecting duplicates among a large number of applications); 2) the task performed by the AI system is intended to improve the result of a previously completed human activity (improving the language used in already drafted documents); 3) the AI system is intended to detect decision-making patterns or deviations from prior decision-making patterns (checking whether the teacher may have deviated from the grading pattern); and 4) the AI system is intended to perform a preparatory task (smart solutions for file handling).

With reference to its personal scope, the AI Act defines different entities like provider, deployer, authorised representative, importer, distributor and operator. For the international arbitration community, the term "deployer" is important. Article 3(4) of the AI Act defines "deployer" as "a natural or legal person, public authority, agency or other body using an AI system under its authority except where the AI system is used in the course of a personal non-professional activity". Recital 13 also states that depending on the type of AI system the use of such a system may affect third parties apart from the deployer. Therefore, arbitrators as natural persons and arbitral institutions as legal persons fall within the personal scope of the AI Act when they use AI systems for professional activity.

Article 26 of the AI Act provides a certain number of regulatory obligations for deployers of high-risk activities, "such as the obligations to (i) take appropriate technical and organizational measures to ensure that the AI systems are used in accordance with their instructions (Article 26(1)), (ii) monitor their operation (Article 26(4)), (iii) assign human oversight to natural persons who have the necessary competence, training, authority and support (Article 26(2)), (iv) ensure the input data is relevant and sufficiently representative (Article 26(4)), and (v) keep the logs automatically generated by the system for a period of at least six months (Article 26(6))" (Scherer, 2024).

Finally, when we evaluate the AI Act's territorial scope, similar to the GDPR, the Brussels effect of the regulation may be observed in Article 2(1). In subparagraph (b), the AI Act stipulates that the regulation applies if the deployer has their location or place of establishment within the EU, while in subparagraph (c) it stipulates the applicability to deployers who have their location or place of establishment in a third country but the output produced by the AI system is used in the EU.

Scherer (2024) argues that the application of Article 2(1)(b) and (c) to international arbitration is not straightforward and explores a few scenarios. The first scenario is with regard to Article 2(1)(b) in which the place of establishment or location of the arbitrator could be their habitual residence. The problem that arises here relates to the constitution of arbitral tribunals in which one arbitrator is from EU territory while the other one or two are not. The second scenario concerns Article 2(1)(c), which provides the potentially extraterritorial effect of the AI Act. As mentioned above, "if AI systems have been used by the arbitral tribunal, the AI system's output has impacted the award, which in turn has legal effects on an EU-based party" (Scherer, 2024).

Considering the aforementioned and Article 2(1)(b) and (c), one may notice that the AI Act potentially has room for significantly greater extraterritorial reach than the GDPR. For the international arbitration community, this would mean the AI Act could be applied even when only one party is located in the EU.

6 Potential Challenges to the Recognition and Enforcement of an Award

Given the categorisation of AI tools used in international arbitration, the risk-based approach of the AI Act and the New York Convention, it should not be surprising that the international arbitration community is becoming even more reluctant towards the use of AI in the arbitral process. So far, the international arbitration community has been hesitant to use AI in arbitral proceedings because AI can bring various challenges related to the risk of bias, lack of empathy, or unreasoned arbitral awards (Waqar, 2022).

Most of these concerns and challenges are due to the final stages of the arbitral process, which is the recognition and enforcement of the awards that emerge. The New York Convention provides specific situations in which an arbitral award may be set aside or unenforceable. Under Article V(2) b of the New York Convention, the recognition and enforcement of an award may be refused if the competent authority in the country where recognition and enforcement is sought finds that the recognition and enforcement would be contrary to the public policy of that country. In the context of the AI Act, if the use of AI in one or more stages of the arbitral process is considered a violation of public policy, such an award could have grounds for refusal according to Article V(2)(b) of the New York Convention. Therefore, if we correlate Article V(2)(b) of the New York Convention with Article 2(1)(b) and (c) of the AI Act and Article 8(a) in Annex III of the AI Act, it is evident that if found that a high-risk AI system was used in the arbitral process, such an award could be unenforceable or even set aside by a national court.

Even though the AI Act was adopted not long ago, the international arbitration community has relied on the Ethical Charter to guide the use of AI in judicial systems and their environment, which provides five core principles: the principle of respect for fundamental rights, the principle of non-discrimination, the principles of quality and security, the principle of transparency, impartiality and fairness, and the principle "under user control" (Berardicurti, 2021; Carrara, 2020). In this sense, the international arbitration community tried to comply with available principles regarding the use of AI to avoid potential challenges to the recognition and enforcement of arbitral awards. Still, with the AI Act bringing stricter rules concerning the use and handling of AI tools to the table, it will be interesting to see how the international arbitration community adapts and changes its approach.

7 The Interaction Between the AI Act and the GDPR

The EU sought to strengthen the protection of the private data of natural persons by introducing the GDPR. In so doing, the EU established basic principles within the GDPR such as lawfulness, fairness and transparency in processing data, purpose limitation in the collection of private data, data minimisation, the accuracy of the collected data, storage limitation, integrity and confidentiality in the processing of data, and accountability for the controller who breaches the aforementioned principles.

After the GDPR came into force, the International Congress and Convention Association (ICCA) and the International Bar Association (the IBA) collaborated to investigate the application of the above-mentioned data protection principles and their potential effects on international arbitration. The task force released a draft of the Roadmap in February 2020 addressing the data protection concerns that could arise in the arbitral process. It is important to note that the Roadmap raised one of the primary concerns with the GDPR; namely, whether any type of arbitration could be excluded from application of the data protection regulations.

In the Coordinated Plan on Artificial Intelligence from 7 December 2018, the European Commission pointed out the need to build up a well-functioning data ecosystem of European data for the further development of AI in Europe. Therefore, trust, data availability and infrastructure were stressed as the main objectives. The European Commission also noted that the GDPR "has established a new global standard with a strong focus on the rights of individuals, reflecting European values, and is an important element of ensuring trust in AI" (EUR-Lex - 52018DC0795 - EN - EUR-Lex, 2018).

Similar to how the GDPR places various obligations on the data controller, the AI Act seeks to place additional obligations upon deployers other than those mentioned above. Article 26 of the AI Act provides additional obligations for the deployers of high-risk AI systems which are interlinked with the GDPR. Here, paragraph (9) sets the obligation for a deployer to carry out a data protection impact assessment pursuant to Article 35 of the GDPR, while paragraph (12) specifies the obligation to cooperate with relevant national competent authorities. Any non-compliance with these obligations will bring administrative fines of up to EUR 15,000,000 or to 3% of its total worldwide annual turnover for the preceding financial year under Article 99(4)(e) of the AI Act.

Recital 39 provides that any processing of biometrics and other private data in the use of AI systems for purposes other than law enforcement is prohibited pursuant to Article 9(1) of the GDPR. On the other hand, Recital 53 sets out the foundation for AI systems that imply profiling within the meaning of Article 4(4) of the GDPR by suggesting that these AI systems should be considered to pose significant risks of harm to the health, safety and fundamental rights of natural persons. Recital 67 overlooks the need for access to high-quality data, which plays a vital role in ensuring the best performance of many AI systems. Considering the GDPR, Recital 67 provides that "data governance and management practices should include, in the case of personal data, transparency about the original purpose of the data collection" so that AI systems perform based on data that is relevant, sufficiently representative and free of errors (*(67) - EU AI Act,* 2024). This is further provided for in the provisions of Article 10(2)(b) of the AI Act.

8 Governance Structures Required to Enforce the AI Act

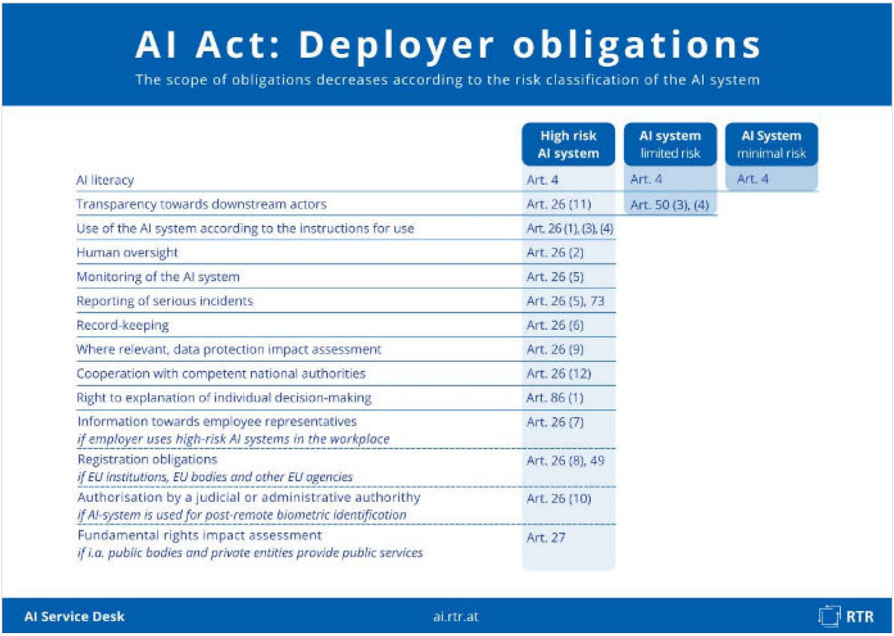
In her article "Cogito ergo (intelligens) sum? Artificial Intelligence and international arbitration: who would set out the rules of the game?", Martina Marganelli asked who would be the best to determine the rules for the use of AI, whether that should be national governments, the business world, international organisations or arbitral institutions. However, the EU was already deep into drafting the AI Act.

Recital 148 articulates the need for a governance framework that allows coordination and support for application of the AI Act on a national level and establish capabilities on the EU level (*(148) - EU AI Act,* 2024). On the latter level, the European AI Office (AI Office) was established by the European Commission on 24 January 2024 (Commission Decision Establishing the European AI Office | Shaping Europe's Digital Future, 2024). Yet, the establishment of the AI Office should not in any way affect the power and competence of national competent authorities within the member states. This is seen in Article 2(3)(d) of the Commission Decision of 24 January 2024 establishing the European Artificial Intelligence Office C(2024)390. Further, it is expected from the AI Office to issue guidance on the AI Act, but in a way that does not duplicate the activities of other relevant EU bodies, offices or agencies. It may be observed from this that member states should facilitate their own national regulators which would work together with the AI Office to coordinate and support application of the AI Act.

> The establishment of the AI Office should not in any way affect the power and competence of national competent authorities within the member states.

From Austria's point of view, a national body designated to coordinate and monitor application of the AI Act is the Artificial Intelligence Service Desk established at Rundfunk und Telekom Regulatory GmbH (RTR) (Artificial Intelligence, 2023). The Service Desk for AI supports implementation of the AI Act but also acts as a point of contact and information for the general public regarding this regulation. In this respect, the Service Desk for AI provides crucial information for citizens about related actors, authorities and bodies, deployer obligations, facts and questions, general information on the AI Act, provider obligations, risk levels of AI models and AI systems, and the time frame (AI Service Desk, 2024).

Figure 1: The infographic provides an overview of the obligations of deployers of AI systems by risk classification



AI Act: Deployer obligations

The scope of obligations decreases according to the risk classification of the AI system

	High risk AI system	AI system limited risk	AI System minimal risk
AI literacy	Art. 4	Art. 4	Art. 4
Transparency towards downstream actors	Art. 26 (11)	Art. 50 (3), (4)	
Use of the AI system according to the instructions for use	Art. 26 (1), (3), (4)		
Human oversight	Art. 26 (2)		
Monitoring of the AI system	Art. 26 (5)		
Reporting of serious incidents	Art. 26 (5), 73		
Record-keeping	Art. 26 (6)		
Where relevant, data protection impact assessment	Art. 26 (9)		
Cooperation with competent national authorities	Art. 26 (12)		
Right to explanation of individual decision-making	Art. 86 (1)		
Information towards employee representatives *if employer uses high-risk AI systems in the workplace*	Art. 26 (7)		
Registration obligations *if EU institutions, EU bodies and other EU agencies*	Art. 26 (8), 49		
Authorisation by a judicial or administrative authority *if AI-system is used for post-remote biometric identification*	Art. 26 (10)		
Fundamental rights impact assessment *if i.a. public bodies and private entities provide public services*	Art. 27		

AI Service Desk ai.rtr.at RTR

Source: AI Act: Deployer Obligations | AI Service Desk, 2024

Moving on to France, the French Data Protection Agency (the Commission nationale de l'informatique et des libertés (CNIL)), as one of the most active European data protection authorities, is expected to assume the central role in the regulation and oversight of AI in France (AI Watch: Global Regulatory Tracker – France | White & Case LLP, 2024). CNIL has already established the Artificial Intelligence Service for the purpose of addressing the relationship between the GDPR and the AI Act. Moreover, CNIL has been actively engaging in investigations and enforcement against actors in the AI industry (Juhan, et al., 2024). Along these lines, CNIL opened an investigation into ChatGPT after having received several complaints and imposed a penalty of EUR 20 million on Clearview AI for its unlawful processing of personal data.

Although France will probably not enact its own national AI regulation, CNIL has published the first set of recommendations for the use of AI to help support relevant actors in their efforts to comply with the GDPR while developing or using AI systems (AI: CNIL Publishes Its First Recommendations on the

Development of Artificial Intelligence Systems, 2024). These recommendations address the aspects of the applicable legal regime, the defining of purpose, the legal classification of the actors, defining the legal basis, carrying out tests and checks in the event of the reusing of data, conducting an impact assessment when necessary, taking data protection into account when designing AI systems, and collecting and managing data.

Germany has yet to designate an authority to act as the national supervisory authority required by the AI Act (Artificial Intelligence 2024 – Germany | Global Practice Guides | Chambers and Partners, 2024). However, currently various ministries of the federal government are participating in international regulatory processes addressing AI (AI Watch: Global Regulatory Tracker – Germany | White & Case LLP, 2024).

These include, but are not limited to: 1) the Federal Ministry for Economic Affairs and Climate Action and the Federal Ministry of Justice which were part of the negotiation process in the drafting of the AI Act (Silicon Saxony, 2024); 2) the Federal Ministry for Digital and Transport engaged in the G7 Hiroshima Process on Generative AI (Hazra & Bisagni, 2024); 3) the Federal Ministry of Education and Research, which published the 2023 AI Action Plan (AI Action Plan Unveiled, 2023); and 4) the Federal Ministry of the Interior and Homeland, which established an AI Advisory Centre (Expo Highlight BMI: BeKI, 2023).

Yet, German data protection authorities (DPAs) are assuming a leading role against companies that are placing AI systems in the German market, at least in the areas where the GDPR can be enforced. Considering that personal data are often used in the training and deployment of AI systems, "DPAs are effectively acting as de facto AI regulators for the time being ⌧ actively working to regulate AI systems and likely to continue playing an increasingly important role in governing AI systems and their handling of personal data" (Artificial Intelligence 2024 – Germany | Global Practice Guides | Chambers and Partners, 2024). While DPAs are presently assuming the leading role as AI regulators, potential candidates for this position include the Federal Network Agency and the Federal Office for Information Security (AI Watch: Global Regulatory Tracker – Germany | White & Case LLP, 2024).

In any case, irrespective of whether member states have appointed national regulatory bodies, many arbitrators and arbitral institutions will abide by the AI Act. This is especially due to the fact that AI systems and tools need data to provide output.

9 Conclusion

There is no doubt that the AI Act will have a big impact on arbitration proceedings and the way AI tools are used going forward. However, the concrete extent of this impact remains quite uncertain as we are still in the early stages of the AI Act's applicability. One thing is sure: the AI Act has a wider extraterritorial effect than the GDPR.

If we consider the above-mentioned four categories of AI tools used in the arbitral process, it is clear that some will fall within the scope of the AI Act as high-risk AI systems. Such AI tools would be those used for selecting suitable experts, counsels and arbitrators since they pose a risk of bias against those less experienced. Others would be AI tools used in the adjudication process, including the tools of 'predictive justice'. These in particular could be completely exempted from further use in the arbitral process as they fall within the scope of Article 8(a) found in Annex III of the AI Act. On the other hand, AI tools used for the efficiency of legal research and for translating, transcribing and summarising evidence could have a longer life within the arbitral process as they could be exempted by Article 6(3)(a) of the AI Act.

Thus far, it seems that one of the ways to address the AI Act's potential risks is for arbitrators and arbitral institutions to disclose the information to all parties involved about the potential use of AI tools in arbitration proceedings, which may have an effect on the arbitral awards. In this way, the parties could voice their opinion or confirm that they abide by the rules of the AI Act. Both arbitrators and arbitral institutions could thereby avoid the grounds for applicability of Article V(2)(b) of the New York Convention. The other and more suitable solution would be to form a taskforce to further evaluate the applicability of the AI Act to arbitration proceedings and produce a Roadmap, similar to when the GDPR was first introduced. In this way, arbitrators and arbitral institutions would have guidelines concerning how to approach the rules the AI Act establishes for them and the arbitral process.

Most of the currently used AI tools have proven themselves useful for lowering the time and costs of arbitration, correlatingly increasing the efficiency of the overall arbitration process. As time, cost and efficiency are key features of international arbitration and given that AI has done revolutionary things in reducing them, it will be interesting to see how the international arbitration community adapts to the newly imposed regulation.

References

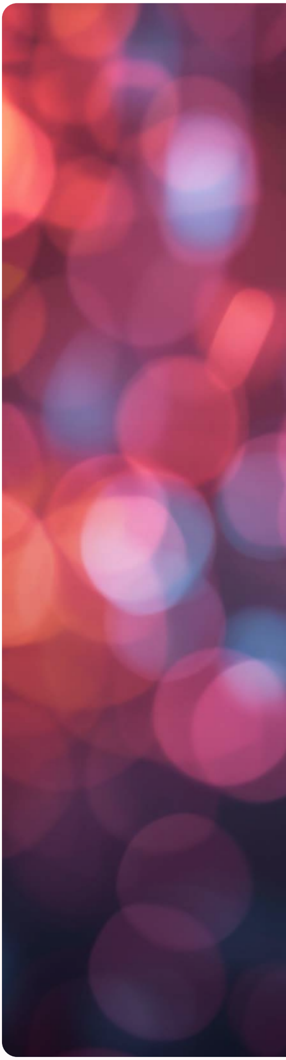
- (53) - EU AI Act. (n.d.). Www.euaiact.com. Retrieved July 18, 2024, from https://www.euaiact.com/recital/53
- (61) - EU AI Act. (2024, May 21). Www.euaiact.com. Retrieved July 18, 2024, from https://www.euaiact.com/recital/61
- (67) - EU AI Act. (2024, May 21). Www.euaiact.com. Retrieved July 19, 2024, from https://www.euaiact.com/recital/67
- (148) - EU AI Act. (2024, May 21). Www.euaiact.com. Retrieved July 19, 2024, from https://www.euaiact.com/recital/148
- About us - Vienna International Arbitral Centre. (2024). Www.viac.eu. Retrieved July 19, 2024, from https://www.viac.eu/en/about-us
- AI Act: Deployer Obligations | AI Service Desk. (2024). RTR. Retrieved July 19, 2024, from https://www.rtr.at/rtr/service/ki-servicestelle/ai-act/Deployer_obligations.en.html
- AI Action Plan Unveiled. (2023, November 8). Www.fz-Juelich.de. Retrieved July 19, 2024, from https://www.fz-juelich.de/en/news/archive/announcements/2023/ai-action-plan-unveiled
- AI Service Desk. (2024). RTR. Retrieved July 19, 2024, from https://www.rtr.at/rtr/service/ki-servicestelle/KI-Servicestelle.en.html
- AI Strategies - Home. (2023). Www.plattform-Lernende-Systeme.de. https://www.plattform-lernende-systeme.de/ai-strategies.html#:~:text=USA-
- AI Strategies and Policies in France. (2024). Oecd.ai. https://oecd.ai/en/dashboards/countries/France
- AI Strategies and Policies in Germany. (2024). Oecd.ai. https://oecd.ai/en/dashboards/countries/Germany
- AI Watch: Global regulatory tracker - France | White & Case LLP. (2024, May 13). Www.whitecase.com. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-france
- AI Watch: Global regulatory tracker - Germany | White & Case LLP. (2024, May 13). Www.whitecase.com. https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-germany
- AI: CNIL publishes its first recommendations on the development of artificial intelligence systems. (2024, June 7). Www.cnil.fr. Retrieved July 19, 2024, from https://www.cnil.fr/en/ai-cnil-publishes-its-first-recommendations-development-artificial-intelligence-systems
- Artificial Intelligence 2024 - Germany | Global Practice Guides | Chambers and Partners. (2024, May 28). Practiceguides.chambers.com. Retrieved July 19, 2024, from https://practiceguides.chambers.com/practice-guides/artificial-intelligence-2024/germany
- Artificial Intelligence. (2023). Digitalaustria.gv.at. https://www.digitalaustria.gv.at/eng/topics/AI.html#:~:text=The%20AI%20Act%20provides%20for,Telekom%20Regulatory%20GmbH%20(RTR).

- Austria AI Strategy Report - European Commission. (2021, September 1). Ai-Watch.ec.europa.eu. Retrieved July 19, 2024, from https://ai-watch.ec.europa.eu/countries/austria/austria-ai-strategy-report_en
- Berardicurti, B. (2021). '25. Artificial Intelligence in International Arbitration: The World is All That is The Case. In C. González-Bueno (Ed). 40 under 40 International Arbitration (2021) (pp. 377 – 392). Dykinson.
- Carrara, C. (2020). Chapter IV: Science and Arbitration, The Impact of Cognitive Science and Artificial Intelligence on Arbitral Proceedings Ethical issues. In C. Klausegger, P. Klein, et al. (Eds). Austrian Yearbook on International Arbitration 2020 (pp. 513-529). Manz'sche Verlags- und Universitätsbuchhandlung 2020.
- Chauhan, A. S. (2020, September 26). Future of AI in Arbitration: The Fine Line Between Fiction and Reality. Kluwer Arbitration Blog. https://arbitrationblog.kluwerarbitration.com/2020/09/26/future-of-ai-in-arbitration-the-fine-line-between-fiction-and-reality/
- CIArb - Brazil. (2024, April 2). What if the arbitrator is replaced by AI? ⊠ | Vis Moot Vienna. YouTube. https://www.youtube.com/watch?v=CrGejdZhpIs&t=1211s
- Commission Decision Establishing the European AI Office | Shaping Europe's digital future. (2024, January 24). Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/library/commission-decision-establishing-european-ai-office
- Coordinated Plan on Artificial Intelligence | Shaping Europe's digital future. (2023, June 30). Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/policies/plan-ai#:~:text=The%20Coordinated%20Plan%20on%20Artificial
- Digital Austria Act - das digitale Arbeitsprogramm der Bundesregierung. (2023). Digitalaustria.gv.at. https://www.digitalaustria.gv.at/Strategien/Digital-Austria-Act---das-digitale-Arbeitsprogramm-der-Bundesregierung.html
- Door Anu Bradford. (2020). The Brussels effect : how the European Union rules the world. Uitgever: New York, Ny Oxford University Press.
- EUR-Lex - 52018DC0237 - EN - EUR-Lex. (n.d.). Eur-Lex.europa.eu. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN
- EUR-Lex - 52018DC0795 - EN - EUR-Lex. (2018). Europa.eu. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52018DC0795
- European Commission I. (2021, April 21). Coordinated plan on artificial intelligence 2021 review | shaping Europe's digital future. Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review
- European Commission II. (2021, April 21). Proposal for a Regulation laying down harmonised rules on artificial intelligence | Shaping Europe's digital future. European Commission. https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence
- European Commission. (2023, January 26). A European approach to Artificial intelligence | Shaping Europe's digital future. Digital-Strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence
- European Commission. (2024, March 6). Regulatory framework on AI | Shaping Europe's digital future. European Commission. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- Expo highlight BMI: BeKI. (2023, October 26). Www.smartcountry.berlin. https://www.smartcountry.berlin/en/newsblog/expo-highlight-bmi-beki.html
- FORWIT. (2023, November 27). Österreichischer Rat für Robotik und Künstliche Intelligenz. Österreichischer Rat Für Robotik Und Künstliche Intelligenz. https://www.acrai.at/

- German Arbitration Institute (DIS): Our history. (2024). Www.disarb.org. Retrieved July 19, 2024, from https://www.disarb.org/en/about-us/history
- Germany AI Strategy Report. (2021, September 1). Ai-Watch.ec.europa.eu. https://ai-watch.ec.europa.eu/countries/germany/germany-ai-strategy-report_en
- Gordon, C.-A. (2020). The Impact of GDPR on International Arbitration – A Practical Guideline (Review of The Impact of GDPR on International Arbitration – A Practical Guideline). Dispute Resolution Journal, 74(4), 25–31. https://go.adr.org/rs/294-SFS-516/images/Gordon%20-%20The%20Impact%20of%20GDPR%20on%20International%20Arbitration.pdf
- Hazra, S., & Bisagni, A. (2024). 2023 G7 Hiroshima Summit Interim Compliance Report (Review of 2023 G7 Hiroshima Summit Interim Compliance Report). G7 Research Group. https://globalgovernanceprogram.org/g7/evaluations/2023compliance-interim/16-2023-G7-interim-compliance-digital-ecosystem-with-trust.pdf
- Juhan, J.-L., Saarinen, M., Martel, D., & Park, A. J. (2024). In Depth: Artificial Intelligence Law | France. Latham & Watkins LLP. https://www.lw.com/en/people/admin/upload/SiteAttachments/Lexology-In-Depth-Artificial-Intelligence-Law-France.pdf
- La stratégie nationale pour l'IA | entreprises.gouv.fr. (2023, October 17). Www.entreprises.gouv.fr. https://www.entreprises.gouv.fr/fr/numerique/enjeux/la-strategie-nationale-pour-l-ia
- OECD AI Policy Observatory Portal I. (2023, November 23). Oecd.ai. Retrieved July 15, 2024, from https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24233
- OECD AI Policy Observatory Portal II. (2024). Oecd.ai. https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-14907
- OECD AI Policy Observatory Portal III. (2024). Oecd.ai. Retrieved July 14, 2024, from https://oecd.ai/en/dashboards/countries/Austria
- OECD AI Policy Observatory Portal IV. (2024, July 5). Oecd.ai. https://oecd.ai/en/dashboards/policy-initiatives/http:%2F%2Faipo.oecd.org%2F2021-data-policyInitiatives-24114
- Our mission, history and values. (2024). ICC - International Chamber of Commerce. https://iccwbo.org/about-icc-2/our-mission-history-and-values/
- Republic of Croatia. (2022, December). Digital Croatia Strategy for the period until 2032 (Review of Digital Croatia Strategy for the period until 2032). https://rdd.gov.hr/UserDocsImages/SDURDD-dokumenti/Strategija_Digitalne_Hrvatske_final_v1_EN.pdf
- Scherer, M. (2022). 'Chapter 39: Artificial Intelligence in Arbitral Decision-Making: The New Enlightenment?'. In C. Bull, L. Malintoppi, et al. (Eds). ICCA Congress Series No. 21 (Edinburgh 2022): Arbitration's Age of Enlightenment? (pp. 683 – 694). ICCA & Kluwer Law International 2023.
- Scherer, M. (2024, May 27). We Need to Talk About … the EU AI Act! Kluwer Arbitration Blog. https://arbitrationblog.kluwerarbitration.com/2024/05/27/we-need-to-talk-about-the-eu-ai-act/
- Silicon Saxony. (2024, February 6). BMWK: Framework for artificial intelligence in the EU in place - AI regulation unanimously approved. Silicon Saxony. https://silicon-saxony.de/en/bmwk-framework-for-artificial-intelligence-in-the-eu-in-place-ai-regulation-unanimously-approved/
- The German Federal Government. (2020). Artificial Intelligence Strategy of the German Federal Government 2020 Update (Review of Artificial Intelligence Strategy of the German Federal Government 2020 Update). https://www.ki-strategie-deutschland.de/files/downloads/Fortschreibung_KI-Strategie_engl.pdf

- The Ultimate Guide to the EU General Data Protection Regulation (GDPR). (2022, March 17). DataGuidance. https://www.dataguidance.com/resource/ultimate-guide-eu-general-data-protection#:~:text=The%20personal%20scope%20of%20the,undertakings%20established%20as%20legal%20persons.
- Villani, C., Bonnet, Y., et al. (2018). For a Meaningful Artificial Intelligence: Towards a French and European Strategy. Conseil national du numérique. https://www.jaist.ac.jp/~bao/AI/OtherAIstrategies/MissionVillani_Report_ENG-VF.pdf
- Waqar, M. (2022). The Use of AI in Arbitral Proceedings (Review of The Use of AI in Arbitral Proceedings). Ohio State Journal on Dispute Resolution, 37(3), 343–366. https://moritzlaw.osu.edu/sites/default/files/2022-08/11-%20Waqar%20Publication%20Final%20343-366.pdf
- Works Constitution Act of 25 September 2001 (Federal Law Gazette I, p. 2518), as amended by Article 6d of the Act of 16 September 2022 (Federal Law Gazette p. 1454)

A liberal future in a united Europe



europeanliberalforum
eurliberalforum
eurliberalforum
European Liberal Forum

This publication was co-financed by the European Parliament. The European Parliament is not responsible for the content of this publication, or for any use that may be made of it.

liberalforum.eu