INSTITUTE OF PUBLIC POLICY

LISBON

# POLICY PAPER 30

## Policy and Regulation of Artificial Intelligence in the AI Act and in Portugal

**André Ilharco** andreilharco@gmail.com

**Steffen Hoernig** steffen.hoernig@novasbe.pt

Policy Papers

The Policy Papers series by the Institute of Public Policy is intended to support public debate with concise works, where public policies are rigorously analyzed and clear recommendations are put forward.

The Authors

André Ilharco is a Researcher at the Institute of Public Policy.

Steffen Hoernig is a Professor of Economics at Nova School of Business and Economics, Universidade Nova de Lisboa.

About the Institute of Public Policy

The Institute of Public Policy is an independent think tank, organized as a non-profit association, whose mission is to contribute to the improvement of analysis and public debate of institutions and public policies in Portugal and Europe, through the creation and dissemination of relevant research.

# Contents

# 1. What is AI and why do we care?

## 1.1 Quick historical overview[1]

Artificial Intelligence (AI) has its roots in the mid-20th century, with key contributions from pioneers such as Alan Turing[2] and John McCarthy[3], among others. Early explorations in the field of AI aimed at constructing intelligent systems capable of emulating human cognitive processes and were primarily directed towards addressing complex problem-solving tasks. The evolution of AI since then has been determined by the world's increasing data storage capabilities and computer processing speed, as well as by their cost. Since early the 2000s, AI developed the capabilities of handwriting and speech recognition (2000), image recognition (2009), reading comprehension (2016) and language understanding (2018), not to mention that at least from 2018 all these capabilities were performed above-human level by AI systems[4].

## 1.2 Definitions and classifications

Defining AI remains a complex undertaking today due to its diverse manifestations. One prominent category encompasses knowledge bases and logical reasoning systems built upon structured data and rule-based logic. These systems aim to emulate human cognition through the utilization of explicit rules and logical inference to derive conclusions and make decisions. Another vital facet of AI is the domain of machine learning, big data, and pattern recognition. This branch focuses on developing algorithms that enable computers to learn from and make predictions or decisions based on data, often in complex and unstructured environments. Additionally, the emergence of generative AI has ushered in a new era of creativity and innovation. Generative AI systems, such as deep learning-based models, possess the capability to produce new content, including text, images, and even entire simulations, by learning the underlying patterns and structures from a given dataset. This

[1] Sections 1 to 5 of this paper comprise our chapter for the PromethEUs paper "Artificial Intelligence: Opportunities, Risks and Regulation" of November 2023, available at https://www.prometheusnetwork.eu.
[2] https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf
[3] https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1911
[4] https://ourworldindata.org/brief-history-of-ai

type of AI has seen remarkable advancements in fields like art generation, content creation, and even the development of highly realistic synthetic media.

## 1.3 Predicted impacts

The main breakthrough AI represents is prediction (Agrawal, Gans, & Goldfarb, 2018)in all of its applications. Concerning benefits, it is predicted that AI will boost productivity, innovation, and efficiency, as AI can process more data more rapidly. This boost will also be supported by its capability to target potential customers more precisely, customize and personalize services and predict customers' needs and preferences. Trade and supply chains will certainly be adapted due to its power to optimize logistics, inventory management, and predictive maintenance in real-time and using large sets of data. AI can perform several tasks of or even fully substitute human labour in many jobs (mainly those with repetitive and routine tasks), ranging from manufacturing, transports, justice to public administration. It is therefore predicted that reskilling will be needed to accommodate the displaced workforce, but also that new jobs will be created for an AI economy.

On other hand, AI systems are opaque, complex, data dependent, and can make autonomous decisions impacting citizens' rights. They have the power to recognize and predict citizens', workers' and customers' whereabouts, preferences, and choices from the data these knowingly or unknowingly provide. Furthermore, interactions with AI systems are becoming increasingly harder to distinguish from those with humans. The capabilities of AI can amplify asymmetries of power and information towards citizens, restricting fundamental rights and liberties, and weakening democracy and the rule of law, if left unregulated and unsupervised.

To cope with these issues, the European Commission proposed a Regulation called the "AI Act" in April 2021 which is currently is in its final stages of negotiation with the European Parliament and the Council of Ministers (the "trilogue"). These negotiations are expected to be concluded by end of 2023, and the Act is expected to formally enter into force in mid-2024.

# 2. Introduction: EU and international background

## 2.1 The road to the AI act

In April 2018, the Commission issued a communication "*Artificial Intelligence for Europe*". This was the first EU document directly paving the way for EU action on AI. The Commission underlined the importance of steering both public and private investment into AI initiatives to keep up with other international actors such as the US and Asia. Investments in these initiatives would be directed towards research and innovation, as well as guaranteeing better data access in all of the EU. Although a Regulation on AI was not yet mentioned, the Commission stated the urgent need for an appropriate ethical and legal framework of AI in the EU, promoting trust and accountability around its development and use.

Just one year later, AI came to be regarded not only as a field demanding coordination and collaboration between Member States, but as a full policy priority for the EU. In von der Leyen's political agenda (von der Leyen, 2019), AI was taken on as an EU policy priority for the period between 2019 and 2024. In 2020, with the *White Paper on Artificial Intelligence*, the Commission mentioned for the first time the idea of a future regulation of AI with a risk-based approach. The latter's purpose is the creation of a proportional regulatory framework that also promotes AI uptake and avoid burdensome regulation on SMEs (in consonance with the EU Data Strategy).

## 2.2 Consistency with EU Charter of Fundamental Rights

AI can affect fundamental rights in the EU, which is why the Commission proposed an AI Act in the first place, designed directly to defend the principles of the EU Charter of Fundamental Rights under AI. As Recital 28 puts it, the EU Charter of Fundamental Rights acts as a compass when determining which AI systems should be classified as high and non-high-risk or even prohibited. The main threats identified by the Commission in the Act affect core rights in the Charter such as the dignity of the human person, an extreme example of which are AI systems of social scoring (Recital 17).

Consistency implies the defence of the rights enshrined in the Charter, but also the proportionality and minimization of some limitation of rights in case of a clash. Restrictions on the use and development of high-risk AI technology may limit the freedom to conduct business (Article 16) or the freedom of art and science (Article 13). Furthermore, its transparency obligations, such as the conformity assessment (Article 19), will affect intellectual property rights (in compliance with the existing EU legal framework on the matter, such as Directive 2016/943).

## 2.3 International context of AI regulation

Council of Europe: The Council of Europe, a Strasbourg-based international forum to promote human rights, democracy and the rule of law, has accompanied technological developments and the rise of AI over the last decade[5]. At present, it is preparing a Convention on Artificial Intelligence that will aim to ensure that the design, development, and application of AI are fully consistent with its values. The preliminary draft shares many concerns with the AI Act proposal, committing its signatories to take the measures needed to protect the three values mentioned above against abuse. For example, in Chapter III, the draft states that parties shall take appropriate measures to preserve the "ability to reach informed decisions free from undue influence [or] manipulation" (art. 9). Additionally, it proposes fundamental principles for the design, development, and application of AI, such as equality and anti-discrimination, privacy and personal data protection, transparency, accountability, among others. Parties must develop measures ensuring the availability of redress and other safeguards, e.g., the right to human review of an AI decision affecting fundamental freedoms or human rights (article 20, 1), as well as the right to be informed that one is being attended by AI rather than by a person (article 20, 2; much like the AI Act, article 52, 1).

USA: Although the US already have laws and regulations on privacy, security and anti-discrimination, there is still no comprehensive legislation on AI. The debate is ongoing, but until very recently it seemed clear that the US government intended any regulation at a federal level on this matter to go slowly and not hinder innovation. For example, while the EU discussing the AI Act, seven big tech firms in AI (Amazon, Anthropic, Google, Inflection, Meta, Microsoft, and OpenAI) were invited to assume public voluntary commitments for

[5] https://www.coe.int/en/web/artificial-intelligence

managing the risks posed by AI while making its development more safe, secure, and transparent. Under these voluntary commitments, companies must ensure the safety of their AI products before releasing them in the markets (which includes both internal and external testing), share information on managing AI risks with stakeholders, and invest in cybersecurity measures to protect sensitive data.

President Biden's Administration has taken other steps on AI. Last year in October, the White House Office of Science and Technology Policy issued a blueprint for an AI Bill of Rights. This is a non-binding document that presents a roadmap for the future development, implementation and use of AI compliant with American citizens' rights. It identifies five core protection principles: safe and effective systems (citizens should be protected from unsafe or ineffective systems); algorithmic discrimination protections (citizens should not face discrimination by algorithms and systems should be used and designed in an equitable way); data privacy (citizens should be protected from abusive data practices via built-in protections and should have agency over how data about them is used); notice and explanation (citizens should know that an automated system is being used and understand how and why it contributes to outcomes that impact them); alternative options (citizens should be able to opt out, where appropriate, and have access to a person who can quickly consider and remedy problems they encounter)[6].

On October 30th, 2023, President Biden presented a new Executive Order (E.O.) on Safe, Secure, and Trustworthy Artificial Intelligence. Diverging from recent meek developments on AI regulation, with this E.O. President Biden places the US on a similar path to the EU. One of the eight core principles indicated by this E.O. to regulate the development of AI in the US is to promote "responsible innovation"[7], a term also used by the Commission in its explanatory memo of the AI Act. Among other policies and political priorities identified in the E.O., President Biden's Administration imposes new standards for AI safety and security. According to the White House, the E.O. "[requires] companies developing any foundation model that poses a serious risk to national security, national economic security, or national public health and safety [to] notify the federal government when training the model and must share the results of all red-team safety tests"[8] (Section 4.2 of the E.O). The E.O. also calls for the US National Institute of Standards to develop "rigorous standards for extensive red-team testing to ensure safety before public release"[9]. These standards shall be applied

[6]https://www.whitehouse.gov/ostp/news-updates/2022/10/04/fact-sheet-biden-harris-administration-announces-key-actions-to-advance-tech-accountability-and-protect-the-rights-of-the-american-public/
[7] https://www.youtube.com/watch?v=fRL1kplm1H4
[8]https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.
[9]https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.

by several US Government Departments to critical infrastructures as well as to address "as well as chemical, biological, radiological, nuclear, and cybersecurity risks"[10]. Additionally, the US Department of Commerce is tasked to develop guidance for content authentication and watermarking to clearly label AI-generated content.

The US political system will still have to digest this E.O. (the Constitutional Court may declare it unconstitutional and strike it down, the Congress may pass legislation that supersedes or nullifies the E.O., among other checks and balance mechanisms that may alter the E.O.'s real and full implementation). But this E.O. brings the US closer to the EU, both by showing a proactive position in terms of AI regulation and also in many principles shared between the EU's AI Act and Biden's E.O.).

United Kingdom: Unlike the EU approach to AI regulation, the UK government will not create any new regulators for AI, nor will it opt for horizontal legislation on AI. Instead, existing regulators were given five core principles[11] (safety, security and robustness; transparency and explainability; fairness; accountability and governance; and contestability and redress) to guide actions on AI. The UK's idea is to leverage the expertise of each regulator within their respective sectors.

The other side of the U.K. approach, as revealed in October 2023, is to promote a global discussion on AI and its risks. The U.K. Government announced the creation of an AI safety body in the UK to evaluate and test new technologies[12]. It will also promote an AI Safety Summit, at Bletchley Park, in early November 2023. Here, the U.K. Prime Minister hopes to bring together international governments, leading AI companies, civil society groups and experts to discuss the risks of AI and how they can be mitigated through internationally coordinated action[13].

These certainly are important initiatives. But UK sectoral regulators are still left to their own devices and to five core principles suggested by the Government in the meantime. Regardless of U.K. Government initiatives, a comparison between UK's vertical and the EU's horizontal approaches to AI regulation as well as their own sense of urgency will provide hints regarding how AI is best regulated: through fluid regulation in each sector, or using a more centralised and concerted approach. Will both work? Will both show timely results? The future will tell.

[10]https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/.
[11] Provided by UK government in policy paper "A pro-innovation approach to AI regulation" Secretary of State for Science, Innovation and Technology (in https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper#:~:text=Our%20framework%20is%20underpinned%20by,Fairness)
[12] https://www.ft.com/content/509012f9-4e08-414c-a97f-dd733b9de6ef.
[13] https://www.gov.uk/government/topical-events/ai-safety-summit-2023/about

China: China is one of the first countries to draft and implement regulations and specific laws on AI. China's authorities (mainly the Cyberspace Administration of China) have directed their attention to specific AI uses, enacting legal instruments such as the Algorithm Recommendation Regulation (ARR, which came into force in March, 2021, and regulates the use of algorithmic recommendation technologies to provide online services in China), the Deep Synthesis Regulation (which came into force on January 10, 2023; deep synthesis technology is commonly referred to as "deepfakes"), the Generative AI Regulation (published on July 13, 2023, and came into force on August 15, 2023), and the Draft Ethical Review Measure (published on April 14, 2023, for public consultation which closed on May 3, 2023, focused on the ethical review of science and technology activities including AI technologies)[14].

China's approach to AI regulation assumes a different form (fragmented with several laws for several AI uses) from the EU approach (horizontal and comprehensive). However, it shares many of the concerns and obligations contained in the current version of the AI Act. For instance, the Deep Synthesis Regulation states that labels be clearly and visibly placed on synthetically generated content[15], a concern shared with the EU (article 53, AI Act[16]). Moreover, the ARR holds operators responsible for establishing a management system that checks "for (…) published information, data security, personal information protections, countering telecommunication network fraud, security assessments and monitoring, and emergency response and handling of security incidents" (ARR, art. 7).

One core difference between the two approaches is the dimension of the political control envisioned in each of the regulatory approaches. For example, in the ARR, China does not only restrict, for example, algorithms from displaying illegal content (Digital Services Act in the EU), but it also demands as an ethical requirement for algorithm operators to adjust their recommendation algorithms in order to adhere to "mainstream values" (ARR, art. 6)[17], compatible with the political and social order, and to prohibit the setting "up [of] algorithmic models that violate laws and regulations, or go against ethics and morals, such as by inducing users to become addicted or spend too much"[18] (ARR, art. 8). Evidently, the Chinese

[14] https://www.lw.com/en/admin/upload/SiteAttachments/Chinas-New-AI-Regulations.pdf
[15]https://www.allenovery.com/en-gb/global/blogs/data-hub/china-brings-into-force-regulations-on-the-administration-of-deep-synthesis-of-internet-technology-addressing-deepfakes-and-similar-technologies
[16] Art. 53, 3: "Users of an AI system that generates or manipulates image, audio or video content that appreciably resembles existing persons, objects, places or other entities or events and would falsely appear to a person to be authentic or truthful ('deep fake'), shall disclose that the content has been artificially generated or manipulated (…)".
[17]https://www.china-briefing.com/news/china-passes-sweeping-recommendation-algorithm-regulations-effect-march-1-2022/
[18] https://www.chinalawtranslate.com/en/algorithms/

government considers the mass surveillance and social ranking of its population a legitimate activity, while these are considered completely inadmissible under the EU framework.

# 3. The AI Act: Proposal and Trilogue Negotiations

## 3.1 The Regulatory Approach in the AI Act

1. Definitions: The AI Act attempts to provide clear definitions of AI systems to ensure proper regulatory application

Due to the fast-changing environment of AI, one of the main concerns in the creation of the AI Act was to provide a clear definition of AI. Article 3, 1 of the Act states that the definitions are made on a high-level technological basis, listed in Annex 1. In this Annex, we can see that these technologies include "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); and statistical approaches, Bayesian estimation, search and optimization methods. The use of these techniques is defined in article 3 as generating outputs oriented to objectives defined by humans. The definition provided by the Commission aims to be as neutral as possible, in order to cover techniques which are not yet developed, and future revisions are foreseen.

The importance of the definition of AI was strongly underlined by several stakeholders consulted in the public consultation launched by the Commission in February 2020, and it is still under discussion in the trilogue. In fact, the European Council's compromise version of December 2022 removes "statistical approaches, Bayesian estimation, search and optimization methods". Both the Council and the Parliament replace the term "software" for "system" and "machine-based system", respectively.

2. Risk-based approach: The Act focuses on high-risk AI systems and imposes specific obligations on their developers and users

The EU approach to AI regulation is based on risk assessments. This means the AI Act regulates AI systems differently according to the risk associated with their design,

development and use. As such, the Commission proposed three levels of risk for AI systems: unacceptable risk; high risk; and non-high risk. Title II states that any AI system of unacceptable risk is prohibited, with a few exceptions for the use of real-time remote biometric identification systems (art. 5, n 2, 3 and 4). In case of infringement, the AI Act establishes administrative fines of up to 30 million euros or 6% of worldwide annual turnover (if the offender is a company). These are the highest values foreseen in the Act.

Title III focuses on high-risk AI systems, both for being a component of a product or a standalone product. Annex III contains the list of high-risk AI systems, which the Commission can amend via delegated acts (art. 7). It mentions AI systems used for the management and operation of critical infrastructure (transport, water, gas, heating, electricity), for biometric identification of natural persons, both real-time and posterior; educational and vocational training; employment, worker management and access to self-employment (for example, AI systems intended to be used for recruitment and selection processes, as well as for making decisions on promotions, task allocation or performance evaluation); law enforcement; migration, asylum and border control management; administration of justice and democratic processes; among others.

According to Ch. 2 of Title III, operators of high-risk AI systems must develop risk management processes capable of identifying and analysing known and foreseeable risks associated with the specific AI system (art. 9, 2a); estimate and evaluate the risks that may emerge when this system is used in accordance with its intended purpose and under conditions of reasonably foreseeable misuse (art. 9, 2b); evaluate other possible arising risks based on the analysis of data gathered from post-market monitoring (art. 9, 2c); and adopt suitable risk management measures appropriate to the sector's harmonized standards and common specifications (art. 9, 2d and 3). The risk management process must be such as to make any residual risk acceptable, and must be communicated to its user (or deployers, using the EP's amended term). Additionally, data sets used to train these AI systems models must be relevant, representative, free of errors and complete (art. 10, 3), and must be subject to appropriate governance and management practices (art. 10, 2). Before being placed on the market, operators of high-risk AI systems must create technical documentation that demonstrates that these systems comply with the AI Act's requirements and must provide national competent authorities with all the necessary information to assess the compliance of the AI system with those requirements (art. 11, 1 and 2). Operators also have record-keeping obligations. These systems must be designed and developed enabling the automatic recording of events. Art. 12, 2, states that these logging capabilities must ensure a level of traceability of the AI system's functioning throughout its lifecycle.

Non-high risk AI systems may be considered either of minimal risk and will not be subject to obligations, or of limited risk if if they interact with natural persons and pose specific risks of impersonation or deception (recital 70). In the latter case, they are subject to the transparency obligations mentioned right below.

3. Transparency and accountability: The Act promotes transparency in AI system behaviour and requires human oversight. Individuals have the right to be informed when interacting with AI.

The first page of the Commission proposal states that "[t]his proposal aims to implement the second objective for the development of an ecosystem of trust by proposing a legal framework for trustworthy AI". In the context of the interaction between citizens and AI technology, trust is based on transparency and accountability. When interacting with high-risk systems, citizens must be capable of interpreting the system's output and use it appropriately, and these systems must be designed and developed to ensure this (art. 13, 1). These systems must give instructions for usage and provide the users with relevant, accessible, and comprehensible information concisely and in a clear form (art. 13, 2). As part of the specific information required in the last paragraph, the Commission obliges high-risk system providers to specify the system's intended purpose and the identity and the contact details of the provider (art. 13, 3). Furthermore, high-risk AI systems must be developed in such a way that they can be effectively overseen by natural persons during the period in which the AI system is in use (art. 14, 1). The human oversight might be performed by the system provider and in some cases by the user (art. 14, 3). Additionally, the Act states that without prejudice to the requirements and obligations for high-risk AI systems (recital 70), those interacting with natural persons must inform them that they are interacting with an AI system (art. 52). AI systems which make use of emotion recognition or biometric categorization, as well as systems generating "deep fakes", must inform the natural person (art. 52, 2) and disclose that the content is artificially created (art. 52, 3), respectively.

4. Data governance and privacy: The Act addresses data governance and privacy concerns, emphasizing privacy-preserving measures and data protection principles.

The AI Act is to be applied without prejudice to the GDPR. There are, however, data protection, privacy, and data governance obligations foreseen in the Act. Training, validation, and testing data sets involved in the development of high-risk AI systems are

subject to data governance and management practices (art. 10). These data sets must be relevant, representative, free of errors and complete (art. 10, 3). The concern for monitoring, identifying and correcting biases in data sets is present in the Act as a "matter of substantial public interest" (recital 44). Thus, article 10, 5 of the Act allows the providers of high-risk systems to process special categories of personal data covered by article 9 of the GDPR. This may be applied in cases where it is strictly necessary to fulfil the above-mentioned purpose, where anonymization may significantly affect it, and "subject to appropriate safeguards for the fundamental rights and freedoms of natural persons, including technical limitations on the re-use and use of state-of-the-art security and privacy-preserving measures". This exception is not to be understood as a free permission for high-risk systems to process special data. In that respect, Recital 41 states that the AI Act shall "not be understood as providing for the legal ground for the processing of personal data, including special categories of personal data, where relevant".

Personal data is also of special relevance in the articles concerning regulatory sandboxes for AI. The AI Act defines these regulatory sandboxes' function as to "provide a controlled environment that facilitates the development, testing and validation of innovative AI systems for a limited time before their placement on the market". If innovative AI systems tested in these sandboxes require the use and processing of personal data, the AI Act determines that this data must "be processed in (…) a functionally separate, isolated, and protected data processing environment" (art. 54, 1, d), must "not be transmitted, transferred or otherwise accessed by other parties" (art. 54, 1, e); and must be "deleted once the participation in the sandbox has terminated or the personal data has reached the end of its retention period" (art. 54, 1, f).

5. Compliance and enforcement: The Act establishes a European Artificial Intelligence Board to oversee implementation and includes penalties for non-compliance

Regarding the enforcement of the AI Act, actors at different levels have different roles. From a national point of view, each Member State must designate one or more national competent authorities for the purpose of supervising the application and implementation of the Act. Among the designated national competent authorities, each Member State shall also designate a national supervisory authority which "shall act as notifying authority and market surveillance authority" (art. 59, 2). Each Member State is responsible for providing adequate financial and human resources to guarantee permanently enough staff and

expertise (in AI and all the areas covered by the Act) for the national competent authorities to fulfil their tasks under the Act (art. 59, 4). National competent authorities are also responsible for the establishment and supervision of regulatory sandboxes (art. 53, 1). Furthermore, Member State shall define penalties for infringements of the Act. These penalties are limited in financial value according to the nature of the infringing AI system (art. 71, 3 and 4). Member States shall also collaborate in the Commission's efforts to "encourage and facilitate the drawing up of codes of conduct intended to foster the voluntary application to [non-high-risk] AI systems (…) of the requirements set out in Title III, Chapter 2" (art. 69, 1).

The Commission has a pivotal role in the implementation of the AI Act. First, the AI Act empowers the Commission to adopt delegated acts to update Annexes I (art. 4), III (art. 7), IV (art. 12, 3), VI and VII (art. 43, 5). Member States are obliged to report annually the status of the financial and human resources of their national competent authorities and an assessment of their adequacy (art. 59, 5). Similarly, national supervisory authorities must report to the Commission the outcomes of relevant market surveillance activities on a regular basis (art. 63, 2). One responsibility attributed to the Commission is to, in collaboration with the Member States, "set up and maintain a EU database containing information concerning (…) high-risk AI systems" (art. 60, 1). This database will be public (art. 60, 3) and controlled exclusively by the Commission (art. 60, 5).

Finally, and crucially for the implementation of the Act, the regulation stipulates the creation of a European Artificial Intelligence Board (EAIB). The Board will be composed of high-level officials of the national supervisory authorities, the European Data Protection Supervisor, and chaired by the Commission (art. 57). Its role is to contribute, assist and advise the Commission and the national supervisory authorities and guarantee a consistent application of the Act (art. 56). Yet, there is still a lot of debate within the EU institutions on whether the EAIB shall assume the form and nature proposed in the Act or if it will need to take on a more permanent and autonomous existence – see the next section.

## 3.2 The Trilogue Negotiations

1. Definitions, again

The versions brought to the negotiation table by the Commission, Council, and European Parliament (trilogue) differ in many points. To start with, there is disagreement about the

definition of AI. In the three proposed versions there is a general consensus on what AI generates: predictions, recommendations, and decisions. But while the Commission proposed a definition of AI formalized in the Annex, both the Council and the Parliament propose a direct definition in the text of the Regulation (which also removes the Commission's power to change this definition). Also, both suggest changing the definition of AI from software to a system. The European Parliament aligned its definition of AI with the OECD's, maybe hoping to increase the international relevance of the AI Act and create another "Brussels effect".

In the end, the main point of contention is what shall be considered AI and therefore what will fall under the AI Act's jurisdiction (take the case, for example, of the inclusion of generative AI in the later versions). How AI is defined in the Act is important, because it must avoid two defects in its application: Under an inflexible definition, the act may not cover crucial AI technology developed in the near future; under a definition that is too wide, it may opaque and invite litigation, thus not be applicable.

## 2. Biometrics and real-time biometric surveillance in public places

The topic of real-time biometric surveillance systems is a protagonist in every discussion of the political and social uses of AI, and the trilogue negotiations are no exception. While both the Commission and the Council included exceptions where this practice could be allowed, such as the search for crime victims of crime or missing children (art. 5, d, i) or terrorism prevention (art. 5, d, ii), the European Parliament banned it entirely. According to the European Parliament's version, remote biometric identification systems shall only be allowed when used ex-post and with prior judicial authorization (art. 5, dd).

The Parliament's vision for privacy and fundamental rights translates into a stronger concern for biometrics and its effects in its version, as compared to the Commission's. While both versions refer to GDPR article 4, the Parliament's version includes definitions of "biometric-based data", "biometric identification" and "biometric verification", which do not exist in the Commission's proposal. Significantly, it expands biometric categorization systems from AI systems able to assign natural persons to specific categories based on biometric data to systems also able to infer categories and attributes from biometric or biometric-based data (art. 3, 1, 35)). Emotion recognition systems can be used not only for the purpose of identifying or inferring emotions or intentions, but also thoughts and states of mind, of natural persons and groups on the basis of their biometric data or biometric-based data (art. 3, 1 34)).

## 3. EAIB vs AI Office

The European Artificial Intelligence Board (EAIB) will act as a coordinating and enforcing force of the Act. Not only will it advise the Commission on the subject, it has many responsibilities of coordination with national supervisory authorities. The form, powers and independence it will have as an institution is crucial to the Act's implementation. This is probably why the European Parliament's version proposes the replacement of the EAIB with an "AI Office". The replacement stems from the idea that the EAIB proposed by the Commission is insufficient, and that the implementation of the AI Act needs a more permanent, independent and resourceful body. According to the EP version of the Act, "[t]he AI Office should have legal personality, should act in full independence, should be responsible for a number of advisory and coordination tasks, including issuing opinions, recommendations, advice or guidance on matters related to the implementation of this Regulation and should be adequately funded and staffed" (art. 76). Contrary to the EAIB which would be chaired by the Commission, the AI Office would be managed by an "executive director (…) responsible for managing the activities of the secretariat of the AI office and for representing the AI office". The Parliament's view seems sufficiently different in terms of which workload to expect from the implementation of the Act, to the point that it even proposes the creation of an AI agency in case an AI Office proves to be insufficient.

## 4. Penalties for Infringements

As mentioned above, the Commission transfers to Member States the powers to define the penalties for infringements of the AI Act. However, the Act also defines maximum penalties in case of prohibited systems (art. 71, 3: "(…) administrative fines of up to 30 000 000 EUR or, if the offender is company, up to 6 % of its total worldwide annual turnover for the preceding financial year"); and for high-risk system infringements connected to data governance (art. 71, 4: "(…) administrative fines of up to 20 000 000 EUR or, if the offender is a company, up to 4 % of its total worldwide annual turnover for the preceding financial year"). Additionally, the Act also stipulates "administrative fines of up to 10 000 000 EUR or, if the offender is a company, up to 2 % of its total worldwide annual turnover for the preceding financial year" for the cases when "incorrect, incomplete or misleading information [is supplied] to notified bodies and national competent authorities in reply to a request" (art.71, 5).

Both the versions of the Council and the Parliament show differences. For instance, the Council agrees with the limit on penalties for non-compliance of prohibited systems but lowers the limit for SMEs and start-ups (3% of of the SME total worldwide annual turnover for the preceding financial year). The protection of SMEs is also present in the cases of misreporting to notified bodies and national competent authorities (lower limit of 1%) and of non-compliance reported on article 71, 4 (limit of 2%). Contrasting with the Council's focus on SMEs, the European Parliament' main criterion to define penalties is the risk level. For prohibited system infringements (article 5), it raises the limit to "40 000 000 EUR or, if the offender is a company, up to 7 % of its total worldwide annual turnover for the preceding financial year". Infringements related to articles 10 and 13 (high risk system obligations of data governance and transparency) are penalized with "administrative fines of up to EUR 20 000 000 or, if the offender is a company, up to 4% of its total worldwide annual turnover for the preceding financial year". Note that in the Parliament's version the penalty limit of 20 million euros is targeted not only to data governance infringements (art. 10), but also to transparency and provision of information to users (art. 13). Any infringement to compliance with any article other than articles 5, 10 and 13 shall be subject to penalties no greater than 10 million euros or 2% of the company's total worldwide annual turnover for the preceding financial year (both values are half of what the Commission proposed for these cases). Finally, the Parliament also cuts by half the penalty limit for cases of misreporting to notified bodies and national competent authorities and does not include any benefits to SMEs.

## 5. Foundational models and General Purpose AI (GPAI)

The fast growth of AI markets and AI capabilities is a challenge for future regulation. The fact that the European Parliament and Council versions were written one year after the Commission's proposal is not without consequences. One of these consequences is the absence of references to "foundational models", "general purpose AI", "generative AI" or any specific AI models in the Commission version. The fact that this absence is important is because, contrary to the conception of AI that justified the use case approach enshrined in the Commission proposal, foundational models and GPAI have a multipurpose nature and capabilities that escape many of the Commission's proposed obligations.

One of the core concerns of both the Council and the Parliament is the downstream regulation, i.e., the regulation across the AI value chain. To make this part of the regulatory discussion clearer, the Council stated in a recent preparatory document for the fourth trilogue session (held on October 24th, 2023)[19] that "certain tailored transparency obligations are necessary to ensure that downstream providers can build AI systems (including general purpose AI systems) on foundation models in a way that is safe and compliant with the AI Act, minimising the risk to violate fundamental rights and safety". This is no minor issue, because, if left unregulated, the AI value chain may suffer from a lack of transparency and information between its constituents. Furthermore, unclear liability and responsibility attributions may also lead smaller enterprises and SMEs to hold back from optimally developing their businesses using specific GPAI: Although these models may be the best for their businesses, it may be too risky to rely on non-transparent and potentially non-complying providers of GPAI[20].

In the above-mentioned document, the Council also defends the application of obligations for all foundational models, assuming both a before-market (such as documenting the model and training process, including the results of internal red teaming, and carrying out and documenting model evaluation in accordance with standardised protocols and tools) and after-market nature (such as providing information and documentation to downstream providers and enabling them to test the foundation models). These obligations are augmented for what the Council defines as "very capable foundation systems"[21] and GPAI used at scale, i.e., with regular external red-teaming, the deployment of a risk assessment and mitigation system, and for the case of very capable foundation models compliance with additional ex-post market controls. Moreover, the Council also mentions the need to introduce obligations to support enforcement of copyright protections as well as obligations to ensure transparency of AI-generated content.

The European Parliament restricts its regulation approach to foundation models (instead of the Council's overarching regulation of GPAI). According to the Parliament's version, providers of foundation models will have to comply with a set of obligations such as data governance measures, performance levels, requirements for energy use, technical documentation, and compliance with certain transparency requirements (art. 28b of the EP version).

[19] https://table.media/europe/wp-content/uploads/sites/9/2023/10/2023-10-17-conseil-ia-mandat-de-negociation-10412dc9fadd4e4fa9b0360960fd13af.pdf

[20] This concern has been reported in Bienert et al., 2023.

[21] According to the Council: "Very capable foundation models should be understood as foundation models whose capabilities go beyond the current state-of-the-art and may not yet be fully understood".

Regardless of the results of the trilogue negotiations, it is likely that foundational models and GPAI will be explicitly regulated. In fact, the EP version states that "foundation models are a new and fast-evolving development in the field of artificial intelligence, it is appropriate for the Commission and the AI Office [or the EAIB] to monitor and periodically assess the legislative and governance framework of such models" (recital 60h). The Council proposes obligations on GPAI that are similar to those for high-risk AI and calls for the Commission to produce an implementing act that further sets out the specific requirements for GPAI (art.4b, 1 of the Council version).

6. Goal definition of the systems and classification of high-risk AI systems

Contrary to the Commission version where high-risk AI systems shall be classified as such if they match the criteria in Annex III, the Council and Parliament provide more specifications for this classification. For example, AI systems classified as high-risk by the Commission are excluded by the Council if "the output of the system is purely accessory (…) and is not therefore likely to lead to a significant risk to the health, safety or fundamental rights" (art. 6, 3). Additionally, the Parliament's version allows for companies developing high-risk AI systems which these do not consider to pose a significant risk (according to the spirit of the Regulation) "to submit a reasoned notification to the national supervisory authority they are not subject to the requirements" (art. 6, 2). After receiving the provider's request, "the national supervisory authority shall review and reply to the [it], directly or via the AI Office, (…) if they deem the AI system to be misclassified". The Parliament's version thus implies the attribution of a filtering power to national supervisory authorities. Both versions of the Council and the Parliament seem to defend a filter-based system for high-risk AI classification. The core idea is to prevent a strict and innovation-repelling classification system, too rigid to comply with in many real-life situations. The downside of a filtering system such as the one proposed by the Parliament or the Council could be legal fragmentation and the creation of high-risk friendly jurisdictions within the EU, which is precisely what this Regulation is trying to fight.

# 4. Implementing the AI Act

## 4.1 Impact and types of regulatory intervention in the EU

According to the EU "New Legislative Framework" of 2008, manufacturers must conduct pre-market controls, ensuring product safety and performance through conformity assessments (CA) according to specific (and essential) requirements defined by law. The idea is based on the understanding that the providers' in-depth understanding of the design and production process makes them the most suitable party to guarantee the conformity of their products with regulatory requirements. Following this logic, in the AI Act the identification of an AI service's risk level and compliance with regulatory requirements is left to the responsibility of the provider[22]. AI firms themselves are therefore called to participate in the process of classifying their systems. As such, while Ch. 2 of the AI Act sets out the legal requirements "for high-risk AI systems in relation to data and data governance, documentation and recording keeping, transparency and provision of information to users, human oversight, robustness, accuracy and security", it leaves the "precise technical solutions to achieve compliance with those requirements"[23] to the discretion of the AI provider. The Act leaves the CA to internal control checks or notified bodies (entities which must be involved as independent third parties (art.33, 4), satisfy specific criteria (art. 30, 1 and 2) and be designated by national notifying authorities (art. 30)) depending on the type of high-risk AI system to be assessed. This includes using internal controls for stand-alone AI systems and employing third-party CAs for AI systems intended to be used as safety components of products regulated under the New Legislative Framework legislation. The Act also explains in detail the CA procedures for each of these types (Ch. 5). CAs must be performed both before the AI system is deployed in EU markets or before putting the AI system into service (art. 19) and (for cases where the providers are distributors or importers) when a high-risk AI system is substantially modified (art. 43, 4). AI product manufacturers also have compliance obligations in specific cases (art. 24).

22 This must not be mistaken with the idea that solely AI systems designers and developers must comply with these requirements. Even if the provider is not the designer/developer of the system, AI providers must guarantee that Ch. 2 requirements are embedded in the system to be compliant. As we have seen in the section 3. B. 5), part of the trilogue discussion is focused on levelling fairly the compliance burden and responsibility all along the AI value chains, especially in the case of foundation models and GPAI.

23 AI Act, Explanatory Memo, 5.2.3, p.13.

According to Recital 64, the envisioned general rule for the CA process of stand-alone high-risk AI systems consists of internal controls and checks when applicable (that is, excluding "AI systems intended to be used for the remote biometric identification of persons"). On the other hand, according to Recital 63, for high-risk AI systems related to products covered by existing Union harmonization legislation, the compliance of those AI systems with the Act should be assessed as part of the conformity assessment already foreseen under that legislation. With this Act, the Commission wants these types of products (for example, machinery, toys, medical devices, etc.) to be subject to the same *ex-ante* and *ex-post* compliance mechanisms of the products of which they are a component, but now ensuring compliance both with sectoral legislation and AI Act requirements.

## 4.2 Subsidiarity structure at the EU level

1. Legal form of "Regulation"

The legislative instrument chosen by the Commission is not random. A Regulation (instead of a directive) harmonizes the regulatory framework inside the EU and avoids legal fragmentation between EU Member States. There are, however, some issues left to each Member State's discretion, such as the definition of penalties for infringements of the AI Act (explained below) or the application of AI for military purposes.

2. European Artificial Intelligence Board

Both the EAIB and an AI Office will impact the AI Act implementation. As we tried to show in the trilogue discussion part, the EU institutions have different views on the workload this entity will have. However, the AI Act clearly defines the powers, constitution and decision process to define rules of procedure of the EAIB if it is to gather the "OK" from all EU institutions. For example, the EAIB adoption of its rules of procedure shall be decided by simple majority (art. 57). Such rules of procedure shall "contain the operational aspects related to the execution of the Board's tasks as listed in Article 58". Of course, the nature of the EAIB tasks is mainly advisory and coordination (recital 76 and art. 58), however, is simple majority the right way to decide? Is this the best way to converge different interests among EU Member States? This may be refuted by the need for the Commission's consent

for the adoption of these rules. Besides all the other supervisory tasks the Act attributes to the Commission, its role in the implementation of the Act is again reinforced.

### 3. Database of High-Risk AI Systems

Title VII of the AI Act establishes the creation of an EU database of registered high-risk AI systems that shall be managed by the Commission (art. 60, 1). The database shall be publicly accessible (art. 60, 3), high-risk AI system providers must register (art. 51), and national competent authorities contribute as well. The Commission will provide all the technical and administrative support to the providers to carry out their system registration. The data available in this database shall contain the data mentioned in AI Act Annex VIII (Art. 60, 2), as well as personal data regarding the "names and contact details of [the] natural persons who are responsible for registering the system and have the legal authority to represent the provider" (art. 60, 4).

### 4. Definition of penalties

As described above, the definition of penalties is a topic that is being discussed among the EU institutions. The three institutions seem to agree that risk is the overarching idea that must define penalties limit (signalling effect), but between them, some points are still to be agreed upon. The Council version wants to include adaptations to penalties to SMEs and the Parliament wants to increase the penalties limits for infringements of transparency and provision of information to users obligations (art. 13).

The AI Act system to define penalties, at the national level, seem to be taken from other previous Acts, such as the Data Governance Act and the Data Act. The problems with it may be the same. On the long run, competition among countries within the EU may arise to soften their AI Act penalties, in order to attract AI providers and developers and thus promote innovation and value to inside of their borders. This possibility may also happen unintentionally since different Member States have different financial and human resources. Of course, the Commission idea to avoid it is to create the EAIB (or an AI Office) to guarantee consistent application and the dialogue between national supervisory authorities. But the same Act requires that each Member State must fully fund and equip their national competent authorities to carry out their tasks (art. 59, 4).

## 4.3 Human oversight

Human oversight is an important requirement of the AI Act for high-risk systems (article 14) and should be a crucial component of the development of human-centric AI. However, the requirement still lacks clarity. The meaning of human oversight for regulating AI is still under discussion. Core questions remain to be answered: What and how is to be supervised? When will the supervision take place? By whom? Take the following example discussed by Enqvist (2023): AI Act article 14, 3 calls not only on the AI designer but also to the AI system user for oversight. At the same time, according to Recital 48: "High-risk AI systems should be designed and developed in such a way that natural persons can oversee their functioning. For this purpose, appropriate human oversight measures should be identified by the provider (…) [S]uch measures should guarantee that the system is subject to in-built operational constraints that cannot be overridden by the system itself and [are] responsive to the human operator". The same recital ends by stating these measures must also guarantee "that the natural persons to whom human oversight has been assigned have the necessary competence, training and authority to carry out that role". These points seem to leave a lot of room for interpretation on the core questions asked above. How can AI systems designers take necessary measures to guarantee that the person "to whom human oversight has been assigned" (recital 48) have the necessary competence, training and authority to perform the oversight of the AI system? Contrarily to what this recital states, article 9, 4 includes a provision which states that after "eliminating or reducing risks related to the use of the high-risk AI system, *due consideration* shall be given to the technical knowledge, experience, education, training to be expected by the user". An obligation of due consideration and guarantees are not substitutes for each other.

The human oversight provisions for high-risk AI systems may also conflict with the GDPR: The latter states that every citizen has the right not to be subject to decisions made by machines (art. 22 of the GDPR). The European Parliament concluded that the AI Act may be declaring a redundant obligation, if not contradicting the GDPR, and proposed the inclusion of a new clause 4a that contains a general principle of human oversight applicable to all AI systems. Its proposal maintained the clause "decisions on specific areas identified in Annex III must be subject to human oversight of at least 2 natural persons", as did the draft of the Council.

## 4.4 Competent National Authorities' actions and identity

Member States can (but need not) create new authorities or designate existing ones as national competent authorities for the purpose of ensuring the application and implementation of the AI Act. Out of the competent authorities, one shall be designated as national supervisory authority, and this one shall serve as the link between the Member State's competent authorities and EU-level authorities connected to the Act (Commission and EAIB/AI Office). By default, the national supervisory authority "shall act as notifying authority and market surveillance authority" unless a Member State communicates to the Commission "organizational and administrative reasons to designate more than one authority" (art. 59, 2 and 3).

In terms of budgetary implications, the implementation of the AI Act will require sufficient technological expertise and human and financial resources which could amount between 1 to 25 FTE per Member State. Much of the implementation costs will be directly influenced by each Member State's current institutional setup and nominations of competent authorities.

No specific authorities or types of regulators are indicated in the AI Act as national supervisory or competent authority. Again, this decision is left to each Member State. However, two facts may lead one to believe that the task of implementing the act will be left to Data Protection Authorities (DPAs). First, the European Data Protection Supervisor is represented in the EAIB (art. 57, 1), consistently with its responsibility to act as the competent authority to supervise Union institutions, agencies and bodies (art. 59, 8). Second, in the EU and globally, data protection authorities have defined policies and taken action regarding AI (e.g., the ban of ChatGPT by Italy's DPA[24], and the French DPA's plan on protecting personal data in the development of AI models[25]).

## 4.5 Impact of compliance for businesses

The limitation of rights in the AI Act are based on what the Commission calls "responsible innovation". This means that the Act restricts some fundamental liberties, such as the freedom to conduct business or the freedom of art and science, in order to prioritize

[24] https://www.euractiv.com/section/artificial-intelligence/news/italian-data-protection-authority-bans-chatgpt-citing-privacy-violations/

[25] https://www.cnil.fr/en/artificial-intelligence-action-plan-cnil

"overriding reasons of public interest", including health, safety, consumer protection, and the safeguarding of fundamental rights. These restrictions are designed to regulate the development and use of AI technology, with a primary focus on preventing and mitigating serious safety risks and likely violations of fundamental rights such as the ones described above. The Commission attempts to keep the limitations proportional to identified risks, ensuring that its constraints are reasonable and necessary, without unduly impeding innovation or damaging the functioning of businesses.

Moreover, the proposal emphasizes the need for increased transparency requirements, which are intended to balance the right to protect intellectual property with the imperative of providing necessary information for individuals to exercise their right to an effective remedy. These transparency measures are also designed to ensure transparency towards supervisory and enforcement authorities in line with their mandates according to this regulation.

According to the impact assessment published by the Commission in April 2021, AI compliance costs will be in the range from 1.6 billion to 3.3 billion euros in 2025 (p. 12) and over 31 billion in the next five years (2023-2028)[26]. To compute this value, the Commission assumed a total of 10% of high-risk AI systems in all of the AI systems landscape. A recent study on the practical perspective of risk classification according to the AI Act identified, from a sample of 100 AI systems, 18% as high-risk and 40% as unclear. Thus the 10% proportion of high-risk AI systems may be way too conservative, and AI compliance costs may be much greater. Compliance costs only (excluding, for example, conformity assessment costs) would represent nearly 17% of total AI investment costs (p.166). It is therefore not surprising that one of the main concerns of the EU institutions while drafting the AI Act is to minimize its burden.

On the one hand, SMEs can expect total compliance costs of up to €400,000 for just one high-risk AI product requiring a quality management system[27]. To provide an idea of the impact, this cost can represent a 40 % reduction in profit for a European business with €10 million turnover (excluding the cost of the AI system itself). On the other hand, SMEs are protected in the Act in several clauses. For example, SMEs will benefit from priority access to AI regulatory sandboxes (art. 55, 1a), will have their "specific interests and needs (…) taken into account [by notified bodies] when setting the fees for conformity assessment under Article 43, reducing those fees proportionately to their size and market size" (art. 55, 2), will receive guidance and advice on the implementation of the Act by national competent authorities (art. 59), and shall have their interests and economic viability taken into account

[26] https://aai.frb.io/assets/files/AI-Act-Risk-Classification-Study-appliedAI-March-2023.pdf
[27] Idem.

by Member States when the latter define rules on penalties (art. 71). Part of the compliance burden may be determined by how heavily other international blocs regulate their AI. Additionally, the Commission's impact assessment also states that the compliance "costs for SMEs could be significantly reduced by sharing systems (e.g. for testing or legal advice)" and that technical and administrative assistance may help to reduce these costs significantly for SMEs (which by article 59 should be guaranteed by national competent authorities).

Businesses other than SMEs are expected to incur in significant costs to comply with the AI Act. Companies operating with AI high-risk systems will have to establish a quality management system, create and manage technical documentation, conduct (or pay for) a conformity assessment (with regular reviews for significant AI system modifications), implement human oversight and continuous monitoring to mitigate potential risks, and ensure compliance with other sectoral laws, the GDPR, and other relevant regulations. If we compare with the GDPR implementation costs (which should be lower, as the AI Act seems to define more requirements than the former), a recent EU study reports that 34% of large companies spent more than 1 million USD to implement the GDPR, while 74% of SMEs more than 100.000 USD[28].

28 European Commission (2021), p. 161.

# 5. Looking forward

## 5.1 Future-proofness of the AI Act and the Commission's powers

All three EU institutions understand that the future-proofness of the AI Act must be guaranteed and allow for later adjustments and implementing acts that adapt the Act to current developments. This comes as no surprise. Other international initiatives on AI regulation are developing in a much less holistic and horizontal fashion. Both the lack of a comprehensive regulation of AI in the US, as well as the UK's (leave regulation to each sector regulators) and China's approach (specific use cases, such as deep synthesis or algorithmic recommendation) show the caution these key international players exercise when regulating AI. Among other factors, one issue is the rapid development of AI technologies and systems that are being produced every day. This speed has already had consequences in the EU legislative process on AI -- the Commission version of the Act did not yet even know of foundation models and GPAI. The Act will for sure require the ability to be updated if it is to endure. The result of leaving important points (such as the list of high-risk AI systems and other compliance specifications) to the annexes is to delegate their updating exclusively to the Commission and its bodies. If this principle is to be maintained , it strengthens the EP's request of establishing an autonomous executive body (such as the AI Office) instead of the EAIB.

## 5.2 Trade-off between legal certainty and restrictions on business models

1. AI providers in the EU

For AI providers in the EU, the AI Act represents two opposing forces. On the one hand, the Act aims to improve legal certainty on requirements and compliance. This is positive – although there are still many points to be decided and clarified in the trilogue, AI may attract a lot of investment, which may be held back due to uncertain liability and compliance rules. On the other hand, the application of rigid, quickly outdated and burdensome regulation may incentivise AI providers to seek opportunities outside the EU instead. The Act and its annexes include several provisions to simplify compliance, specifying its requirements, as

well as providing regulatory sandboxes (the possibility of testing AI systems outside these sandboxes is being discussed in the trilogue). In the long term, because the AI Act is globally the first horizontal and comprehensive AI regulation, the size of this deterrence effect will depend on the virtues and vices of competing regulations in other international blocs, as well as on the EU's and Member States' incentives and policies supporting AI. The future will tell.

2. AI users in the EU

From the point of view of AI users, the AI Act will promote their fundamental rights and liberties, because the Act guarantees compliance with EU requirements for the AI systems they use. While there are benefits (such as mandatory transparency and information provision requirements for most AI systems), users that deploy AI systems (and therefore must comply with the Act) have now a clearer idea of their share of responsibility in the AI value chain.

## 5.3 Brussels effect, or common regulatory approach US-EU-China?

With the AI Act, the Commission hopes to repeat the regulatory success of setting the agenda and the rules with the GDPR: the "Brussels effect". Will there be another one? The benefits of a Brussels effect would be clear: rules in international markets harmonized with those in the EU, and promotion of EU principles and human rights worldwide in what concerns AI. This would be a big deal, due to AI's potential for both economic growth and political misuse.

There are two reasons that point towards a future Brussels effect in AI. First, both GDPR and the AI Act have extra-European scope, that is, both force companies outside the EU and international actors to comply with EU principles in order to participate in EU markets. Second, although other jurisdictions are already starting to regulate aspects of AI, the AI Act is the first comprehensive AI regulation. This is important because, in the case of prolonged regulatory inaction of other actors such as China or the US, the AI Act may start to impact and even transform companies outside of the EU and their business models, increasing the opportunity costs of other actors' regulatory initiatives if they contradict the EU's AI Act.

There are however several factors that reduce the likelihood of a renewed Brussels effect. First, in contrast with GDPR, the AI Act may imply more burdensome compliance costs, which can deter international AI companies from acting and investing in EU markets. Second, the majority of big AI companies are found outside the EU, such as in the UK, China and the US, whose governments also seem to be quite a bit more receptive to their lobbying against tighter regulation. This leverage strongly increases the likelihood of a London/Beijing/Washington effect rather than a Brussels effect. It is important to note that the different blocs still have different ideas on how (and when) to regulate AI. Recently, the UK Prime Minister asked in a public address regarding AI regulation: "How can we write laws that make sense for something that we don't yet fully understand?"[29]. Less than a week later, in the context of President Biden's Executive Order on AI Safety, White House Chief of Staff, Jeff Ziently, declared publicly "given the pace of this technology [AI], we can't move in normal government or private-sector pace, we have to move fast, really fast – ideally faster than the technology itself"[30]. These two statements reflect two different positions between the US and the UK but also reflect the uncertainty regarding the shape an international regulatory framework for AI will take and who will influence what.

Yet, regardless of the epicentre of regulatory influence, converging interests could lead to converging international regulatory frameworks. This convergence may be delayed or blocked by three factors. First, competition to attract innovative AI companies and technologies. Each bloc may feel compelled to regulate little as possible and to promote public policies and investments to stimulate the development of AI within their jurisdiction. Second, the role of AI for military purposes. The absence of compliance requirements for military purposes and the military industry in the EU AI Act is no chance omission. Finally, as we can see if we compare the EU approach to China's, AI regulation in each bloc is designed to protect different social and political principles (individual liberty and rights, or social order and patriotism, apart from innovative activity). There may be little room for agreements that can bridge these fundamental differences.

However, these three factors have potential arguments against them. First, these countries share economic interests and benefit from the harmonization of the markets, opening their economies to new customers. Second, international convergence of the AI regulatory landscape will increase transparency and control over foreign AI systems, which will contribute to a greater control and better implementation of local regulations on AI.

[29] https://www.ft.com/content/509012f9-4e08-414c-a97f-dd733b9de6ef

[30] https://edition.cnn.com/2023/10/30/politics/white-house-tackles-artificial-intelligence-with-new-executive-order/index.html

In conclusion, international regulatory frameworks on AI may converge, similar to what happened with the GDPR – but this time convergence likely will not be centred around Brussels, EU legal requirements, or even the EU's social and political principles.

# 6. Artificial Intelligence in Portugal: Rules, Rights, and Strategies

## 6.1 The Portuguese Constitution

The Constitution of the Portuguese Republic was approved on April 2nd, 1976, and entered into force on April 25th, 1976. The initial text has been revised seven times since then, and as of now there is no article referring to artificial intelligence specifically – obviously, concerns about its use are much too recent. The text includes, however, since its first version, the right to protection of personal data and limitations on the use of information technology (article 35).[31]

According to the present form of article 35, written in 1997, Portuguese citizens have the right to access, update and correct their computerized data, as well as to know the end to which these will be used (art. 35, 1). Protection measures related to the authorized treatment, transmission, and use of personal data are to be set out in law (later defined in the GDPR) and overseen by an independent entity (implemented as the National Commission for Data Protection [*Comissão Nacional de Proteção de Dados*]; art. 35, 2). Also, information technology is restricted from handling certain sensitive data (such as political convictions, party or union affiliation, and religious faith, among others) without the explicit consent of its owner (art. 35, 3). Unauthorized access to others' data is prohibited, as is assigning citizens a unique national number (art. 35, 4 and 5). Public network access is guaranteed, with the law regulating cross-border data flow (art. 35, 6). Finally, manual files enjoy the same protection as automated data, in line with the law (art. 35, 7). These protections of personal data are in line with the prohibition of certain uses of AI in the EU AI Act, such as the ban on AI technologies for social scoring of natural persons (Recital 17; art. 5, c). The same is true of the limits to and protection against surveillance in the special regime created with the classification of high-risk AI systems (Recital 18; Title III).

31 See https://diariodarepublica.pt/dr/legislacao-consolidada/decreto-aprovacao-constituicao/1976-34520775 (choose article 35 in the dropdown menu, then select "Ver diferenças entre verso¨es deste artigo" to see how it was adapted over time).

## 6.2 The Strategic Framework

### Public Initiatives

The Portuguese national strategy for AI was developed in the context of the "Iniciativa Nacional Compete^ncias Digitais e.2030" (INCoDe.2030)[32], a programme to foster investment in digital skills, combining the efforts of the Portuguese government, businesses, academia, research institutions, and public administration. The launch of this initiative was first discussed in late 2016 and took place in April 2017. The INCoDe.2030 highlights five lines of action: inclusion, education, qualification for employment, specialisation, and research. Later in 2017, INCoDe. 2030 included specific sessions on AI in its First National Forum on Digital Skills. However, the Portuguese 2030 AI strategy as such only started to be discussed in late 2018 and was launched in February 2019,[33] after a series of preparatory meetings, public consultations with key stakeholders, and public discussions with AI experts. AI is one of INCoDe's five national strategies, the others being advanced computing, data, Web 3.0, and cyberspace security. In June 2023, the Portuguese Secretary of State of Digitalization and Administrative Modernization started a process of revision of INCoDe.2030 which will include an update of the Portuguese AI Strategy. This process started by nominating three members from academia (from the areas of Computation, Economics and Law) to lead the three strategic axes of the strategy (AI, Web 3.0, and Data), as well as a new general supervisor of the whole process.

Already before the launch of the National AI Strategy 2030 in 2019, INCoDe.2030 promoted AI initiatives during 2018. Much of those had to do with European initiatives on the matter, such as the Portuguese participation in the preparation of the "European AI declaration" with the Commission's DG Connect (January 2018) and the signing of the European AI declaration (April 2018) or were focused on the application of AI technology and knowledge to the public administration in Portugal, namely developing and funding new research activities and further developing its AI competences.

Since the beginning of the definition of INCoDe.2030, the AI ecosystem has been propelled by State initiatives in collaboration with academia and private actors. First, as promotion of R&D activities, the Portuguese Foundation for Science and Technology (*Fundação para a Ciência e a Tecnologia*, FCT) launched in March 2018 a call for R&D projects named "FCT´s Mobilising programme to foster AI in the public administration", which resulted in the presentation in October of that year of 19 projects funded by the FCT.

32 https://www.incode2030.gov.pt/

33 https://bussola.gov.pt/Estratgias%20e%20Orientaes/Estrat%C3%A9gia%20Nacional%20de%20Intelig%C3%AAncia%20Artificial.pdf

The Portuguese AI Ecosystem has also been strongly influenced by two initiatives: the Collaborative Laboratories (CoLABs) and the Digital Innovation Hubs.

The CoLABs can assume the form of a company or a non-profit organization and must include one company and one R&I (Research and Innovation) institution recognized by the FCT or a State Laboratory. The main goal of the CoLABs is to create qualified and scientific employment in Portugal through the collaborative adoption of R&I agendas by the actors within each CoLAB partnership (which can be the State, social or private actors). The three types of actors are called to fund each CoLAB project. Until the end of 2020, 68.6 million euros were given through national and European funds to CoLAB partnerships, which resulted in the creation of 466 jobs (33% of which were occupied by PhD holders). The CoLABs focus on the areas of health and ageing, green hydrogen, thermal waters, data science, sustainable aquaculture, and tourism.

The Digital Innovation Hubs (DIH) are part of the Action Plan for Digital Transition launched by the Portuguese Ministry of Economics. The DIH are consortiums that integrate several entities, from CoLABs to clusters, including Technological Interface Centers (CIT), academia, civil and private associations, and public or private entities. The idea is to create a supportive network for Portuguese SMEs, companies, and the public administration to help and accelerate their process of digital transformation. Currently, there are 17 DIH in Portugal, out of which thirteen were selected in 2023 to receive 68 million euros by the Recovery and Resilience Plan and to integrate the European Network of DIH. According to the Ministry, these 13 DIH will be responsible for helping with the digital transition of 4977 companies in Portugal by the end of September 2025.

Regarding the most recent applications of AI to the Portuguese Government, through its ministries and administration, two applications stand out. The first was developed by the Portuguese Ministry of Justice in collaboration with the company Genesis.Studio with the support of Microsoft and is called the *Practical Guide to Justice* (GPJ)[34]. By making use of GPT 3.5, the Portuguese Justice website (*Portal da Justiça*) that was created for citizens to access information on their judicial activity now displays a generative AI chatbot to help citizens with Portuguese law. In its initial phase (the GPJ was launched in March 2023), the chatbot is helping citizens search for information about family law, namely on the topics of marriage and divorce. The case of the GPJ is important for Portugal's public services because it may serve as an example of digitalization to many others. In fact, for the next phases, the GPJ is expected to cover more of the interaction between the State and citizens (such as by

[34] https://tek.sapo.pt/noticias/internet/artigos/guia-pratico-da-justica-suportado-por-gpt-ja-da-informacoes-sobre- casamentos-e-divorcios

providing information on how to register a birth, create a company, or obtain an online criminal record or an electronic judicial certificate).

Indeed, the GPJ example was followed, just two months after its launch, by the second example of the application of AI: the *ePortugal* virtual assistant (chatbot). In May 2023, the Portuguese Agency for Administrative Modernization (AMA) adopt a chatbot to help citizens with their digital mobile key (*Chave Móvel Digital,* CMD). The CMD allows Portuguese citizens to access several Government web services, make requests regarding their public information, as well as provide a State-certified signature on digital documents. Using GPT 3.5, the *ePortugal* chatbot is designed to answer citizens' questions about the CMD, and the expectation of its creators (AMA in collaboration with Microsoft, DareData Engineering and Defined.ai) is that it will also be able in the future to answer questions on other topics, such as the ID cards or driving licenses[35].

## Private Initiatives

Portugal's AI landscape is also being transformed by private initiatives. These initiatives tend to take on the form of consortium, and some have been granted substantial financial support by the Portuguese Government. The first initiative is Accelerat.ai and is a partnership between Microsoft, DareData Engineering that is led by Defined.ai. Accelerat.ai is working on solutions to automate customer support with AI technologies developed in European Portuguese (instead of English). It has received 34.5 million euros in PRR funds and is planned to be active until December 2025. One of the outputs developed within the Accelerat.ai partnership is Albertina PT, an AI project developed by researchers at the Universities of Lisbon and Porto. Its goal is to create the first major generative AI model in European Portuguese, in open source and with universal access free of charge, also available in Brazilian Portuguese.

A second private initiative is the Center for Responsible AI (CRAI). The CRAI aggregates 10 start-ups (such as Unbabel, Feedzai and others), 8 research centres, 1 law firm, 2 unicorns, and 5 companies as well as many Portuguese personalities connected to science and technology. The Center offers AI products aimed at contributing to a fairer society, developing AI systems that are transparent, eco-friendly, explainable, and trustworthy. Along with its products, the CRAI also produces research in the areas of energy-efficient and sustainable AI; privacy-preserving AI systems; transparent, fair, and explainable AI;

[35] https://tek.sapo.pt/noticias/computadores/artigos/assistente-virtual-do-eportugal-pode-responder-a-novos-temas-e-permitir-completar-servicos

language technologies and embodied Human-AI interaction; and multilingual and contextualized conversational AI. The Center has received a State PRR fund of 78M euros.

## 6.3 Portuguese Charter on Human Rights in the Digital Era [Carta Portuguesa de Direitos Humanos na Era Digital]

In May 2021, Portugal approved a Charter on Human Rights in the Digital Era.[36] The Charter was proposed following the EC Communication *Action Plan against Disinformation* (JOIN(2018) 36 final). The main topics of the 23 articles of the Charter are the fight against discriminatory practices and situations to limit or prohibit citizens' access to the Internet (arts. 3 and 5), freedom of expression, content creation, and free association rights *online* (arts. 4, 7 and 16), privacy in the digital world (art. 8), internet neutrality (art. 10), right to identity and the State's obligation to protect it (art. 12), the right to be forgotten (art. 13), digital platforms and cybersecurity (arts. 14 and 15), digital will (art. 18), the right to protection against abusive geolocation (art. 17), among others. Also, the Charter delegates to the Portuguese State the task of fomenting digital skills at every age, while it also declares rights *vis-à-vis* the public administration (art. 19).

Many of these points can be related to, or may even collide with, more recent European initiatives, such as the AI Act and the GDPR. One of these touchpoints resides in the Charter's article 13 on the "use of artificial intelligence [AI] and robots". Article 13, 1 states that "the use of [AI] must be guided by *respect for fundamental rights*, ensuring a fair balance between the *principles of explainability*, *security*, *transparency,* and *responsibility*, which meets the circumstances of each specific case and establishes processes aimed at avoiding any prejudices and forms of discrimination". As we have seen in previous sections, the AI Act does include several provisions on the explainability, transparency and security of the AI systems deployed in the EU. The Act will thus determine the "fair balance" of these principles in "each specific case" in each specific risk level the system represents (as identified by the AI Act).

The Charter's article 13, 2 may be more problematic.[37] It states that "decisions with a significant impact on the recipients that are taken through the use of algorithms must be communicated to interested parties, being subject to appeal and auditable, under the terms provided for by law". Many questions arise here. The first is that it is hard to think of a

36 https://www.parlamento.pt/Legislacao/Paginas/Educacao_Carta-Portuguesa-de-Direitos-Humanos-na-Era-Digital.aspx

37 Charter art. 13, 2: "As deciso˜es com impacto significativo na esfera dos destinata´rios que sejam tomadas mediante o uso de algoritmos devem ser comunicadas aos interessados, sendo suscetí´veis de recurso e audita´veis, nos termos previstos na lei."

"decision with a significant impact on the recipients" that is legally *prohibited* to be appealed in democratic countries under the rule of Law. This clause thus seems to be unnecessary. Even so, it may be interpreted as creating a legal obligation for the decisionmaker to answer the citizen's appeal and explain the decision if asked to do so.

A second question is related to the application of art. 13, 2 and how it communicates with GDPR article 22, 1 (which states that "the data subject shall have the right not to be subject to a decision based solely on automated processing (…) which produces legal effects concerning him or her or similarly significantly affects him or her.) or the AI Act article 14 (on human oversight). At first glance, art. 13, 2 clause seems to reinforce GDPR article 22, allowing citizens not only not to be subject to decisions solely based on algorithms but also legally protecting them with the creation of the obligation of the decision-taker to inform them about the use of algorithms on decisions affecting them significantly, while also creating a right for them to appeal and audit these same decisions. But again, in decisions taken by algorithms under human oversight, that is, with the indirect participation of humans in the deliberation, is the decisionmaker free of article 13 obligations? What happens in scenarios where algorithms are used to inform or support the decisionmaker? Where is the threshold? Shall this article be interpreted as stating that every decision that makes use of an algorithm must be auditable and may be subject to appeal (something the AI Act seems to facilitate) or shall we consider under this clause only the decisions taken *solely* based on the results of an algorithm?

The fact is that some points of the Portuguese Charter on Human Rights in the Digital Era were revoked only two months after being published. Its universal nature (Human Rights) also does not seem to be a political matter to be decided at national level. Even if it assumes more of a programmatic nature, the need for revision will arise soon, with special regard to AI, after the AI Act enters into force in 2024.

## 6.4 Ibero-American Charter of Principles and Rights in digital environments [Carta Ibero-americana de Princípios e Direitos Digitais]

During the recent Ibero-American Summit (XXVIII) of the Heads of State and Government, in Santo Domingo, Dominican Republic, on March 25th, 2023, Portugal cosigned the Ibero-American Charter of Principles and Rights in digital environments.[38] Although, as the Charter mentions in its introduction, this document does not have a binding

[38]https://www.segib.org/pt-br/?document=carta-ibero-ameriana-de-principios-e-direitos-em-entornos-digitais

nature (its content has no direct application in national judicial systems), it declares a set of principles each Government commits to promote both in its public policies and legislative initiatives and also in its relationship with national stakeholders, such as businesses and academia.

Among the long list of principles affirmed in the Charter, the centrality of the human person in digital environments is crucial. While acknowledging that technological development often progresses faster than normative or regulatory processes (which is clearly visible in the legislative process of the AI Act in the EU), the Ibero-American States committed themselves to promoting an information society centred on the individual, granting universal, non-discriminatory, and equitable access to digital infrastructures and ICT. The Charter's principles and compromises extend to digital inclusion and connectivity (with special attention to vulnerable groups, gender and age); privacy, trust, data security and cybersecurity (where it is affirmed that digital information systems must be built from the start with appropriate safety measures to guarantee the integrity, confidentiality, availability, resilience and authenticity of the data used in them, as well as of the services they provide; also, the Charter enunciates a collective effort to promote the Ibero-American cooperation for data interoperability).

The Charter's signatories also commit to full access to education, culture and health in inclusive and safe digital environments; special attention to children and adolescents (namely by guaranteeing the application in digital environments of the dispositions of the Convention on the Rights of the Child; and also in fighting the display of violent content, bullying, and any other humiliating practices against children and adolescents); the defence of social, economic and political participation in fair and sustainable digital environments (the need for State intervention and regulation on digital environments is reaffirmed; the role of each State must be to fight disinformation and to protect fundamental rights and freedoms, such as to free expression, to be informed, to have open access to public information).

Concerning public administration and the economy, commitments include the digitalization of the public administration, protection of citizens' personal data and the increase of the access and reach of public services; a fair, inclusive and safe digital economy (the process of digital transformation must promote sustainability and ensure the application of labour rights; it must guarantee equitable growth and integration of all national regions); and emerging technologies that do not renounce the centrality of people (the signatories state that the Charter will be revised and updated in the future due to the unforeseeable development of emerging technologies, among which AI and UNESCO's *Recommendation on the Ethics of Artificial Intelligence* are mentioned); and finally to

guarantee cooperation and mutual assistance between Ibero-American countries in the efforts of digital transformation.

## 6.5 Lisbon Declaration: Digital Democracy with a purpose [Declaração de Lisboa sobre direitos digitais]

European Union presidencies have recently been keen on the discussion of which principles will guide European societies in a digital world. This is what motivated the 2017 Tallinn Declaration on eGovernment and the 2020 Berlin Declaration on Digital Society and Value-based Digital Government. These declarations were produced due to the action of the Estonian and German presidencies of the EU. More recently, during the Portuguese presidency, the Portuguese government also promoted the discussion on these matters and about the "2030 Digital Compass: The European Way for the Digital Decade",[39] resulting in the publication of the 2021 Lisbon Declaration.[40]

The title of the Lisbon Declaration is "Digital Democracy with a purpose". All EU countries agreed on the principles inscribed in the declaration, related to digital identity, privacy, data protection and cybersecurity, access to, use and neutrality of the internet, use of artificial intelligence, freedom of expression and information, freedom of assembly and association, child protection, care and freedom of expression, digital education, digital platforms, digital public services, copyright and other intellectual property rights, digital legacy, and effective remedy and access to justice.

39 https://eur-lex.europa.eu/legal-content/en/TXT/?uri=CELEX%3A52021DC0118

40 https://www.2021portugal.eu/pt/noticias/declaracao-de-lisboa-sobre-direitos-digitais-e-pontape-de-saida-para-uma-carta-internacional

# 7. Conclusion: Portugal is changing its approach to AI, but will AI change Portugal?

As we have seen with the public and private initiatives, the main AI projects being developed in Portugal are focused on the development and opening-up of AI codes and systems, as well as on the automation of customer support services (both public and private). This is important for Portuguese-speaking citizens, as well as for companies operating in Portuguese-speaking countries. The private sector is receiving support in public investment to find AI solutions. However, recent international reports such as the Global Innovation Index 2023[41] state that fragilities of the Portuguese economy may slow growth and innovation powered by AI. This report indicates as some of Portugal's weaknesses its policies for doing business, the level of firms offering formal training, as well as fragilities in the impact of knowledge in firms (especially related to AI is the low performance on the indicator of labour productivity growth); thus the Government's urgency to reinforce digital access to public services, namely through its last investments in AI chatbots on Government websites such as GPJ and *ePortugal*. Still, according to recent reports, Portugal is keeping up with the EU. In fact, according to the European Commission's Digital Economy and Society Index 2022[42], Portugal's digital public services for citizens are above the EU level. Portugal's digital public services for businesses are at EU average level. It is important to remember that both the GPJ and *ePortugal* projects were launched in early 2023, while the report provides values for the year of 2021.

The Portuguese Government's most recent investments in the use of AI to increase and improve the access of citizens and businesses to public services are in line with most of its international commitments, such as the Lisbon Declaration or the Ibero-American Charter of Principles and Rights in digital environments. As we tried to make clear in this paper, the EU AI Act will most likely have a profound impact on European businesses, potentially entailing heavy compliance costs for businesses. As such, although keeping up and leading in several indicators such as digital access to public services in the EU may be good news, the upcoming impact of the AI Act may require a different approach, namely by promoting and strengthening Portuguese businesses and their capacity to apply knowledge both about and created by AI and maximize its impact.

[41] https://www.wipo.int/edocs/pubdocs/en/wipo-pub-2000-2023-en-main-report-global-innovation-index-2023-16th-edition.pdf
[42] https://digital-strategy.ec.europa.eu/en/policies/desi-portugal

This change may have accelerated in June 2023 with the beginning of the revision of INCoDe.2030 and therefore the Portuguese 2030 AI Strategy. The parliamentary elections of 2024 will substitute a government that has been in office for almost 9 years. The perspectives of change and the need for most of the political parties to reinvent themselves, to present new ideas and a new vision to the Portuguese society and economy, may also be good reasons to believe in a political shift of priorities and awareness of both risks and opportunities provided by AI.

# IPP POLICY PAPER 30

*Policy and Regulation of Artificial Intelligence in the AI Act and in Portugal*

Authors: André Ilharco and Steffen Hoernig

ISSN: 2183-9360

January 2024

INSTITUTE OF PUBLIC POLICY

LISBON

Institute of Public Policy Lisbon – Rua Miguel Lupi 20, 1249-078 Lisboa PORTUGAL
www.ipp-jcs.org – email: admin@ipp-jcs.org – tel.: +351 213 925 986 – NIF: 510654320