The AI Act: the evolution of ”trustworthy AI” from policy documents to mandatory regulation

Mélanie Gornet

The AI Act: the evolution of "trustworthy AI" from policy documents to mandatory regulation

Mélanie Gornet
Télécom Paris – Institut Polytechnique de Paris
melanie.gornet@telecom-paris.fr

Abstract

What with the dangers of artificial intelligence for individuals and society, and the rapid evolution of these technologies, Europe has decided to take the lead by imposing strict requirements for the placing on the market of "AI systems". This new European law, adopted in the summer of 2024, is better know as "the AI Act". The AI Act is based on a hierarchy of risks, where riskier systems will be subject to stricter obligations. While the AI Act is not the first law in Europe to be based on risk – the General Data Protection Regulation (GDPR) and subsequent laws on digital technologies have already started this trend – it is the first to take it to such a level. But the AI Act also draws on the concept of "trustworthy AI", a term coined by policy documents that preceded it, and according to which AI must notably be ethical and technically robust.

In this work, we retrace the story of the AI Act, in order to understand the origin of its main concepts and structure. We also take a look at the final version of the text, its hierarchy of AI systems and the corresponding obligations, as well as the governance ecosystem it puts in place to ensure that these rules are properly implemented. The picture we draw shows a regulation that is quite unique in the European legal landscape, despite its many roots and inspirations.

1 Introduction

The European AI Act is the first mandatory framework adopted for AI in the world. At the time of writing, the final text has been published in the Official Journal of the European Union (OJEU) (European Parliament and Council, 2024) and entered into force on 1 August 2024, although some requirements will apply later. This official endorsement follows months of negotiations between the three European institutions, the Commission, the Council and the Parliament, during which the latter two each suggested amendments to the text initially proposed. The European Union prides itself in this achievement, which took years in the making. Indeed, while the first proposal of the AI Act was published in 2021, the idea of developing a mandatory framework for AI in Europe is much older, being mentioned in policy documents dating back to 2017. What makes the AI Act's approach so different from other European legislation?

We start in Section 2 by recounting the story of the AI Act, from the first discussions on creation of a legal status for robots, through the recurring discourse on AI "trustworthiness" and "risk", to the final process of adoption of the text we know today. In Section 3, we show that the AI Act is not alone in the legal landscape of digital technologies in Europe and builds on previous frameworks for data protection, digital platforms, product safety, product liability and so on. In Section 4, we analyse one of the legal texts that strongly inspired the AI Act: the General Data Protection Regulation (GDPR). We show that the GDPR laid the foundations for a risk-based regulation in which technical standards would play a prominent role, although it did not go as far

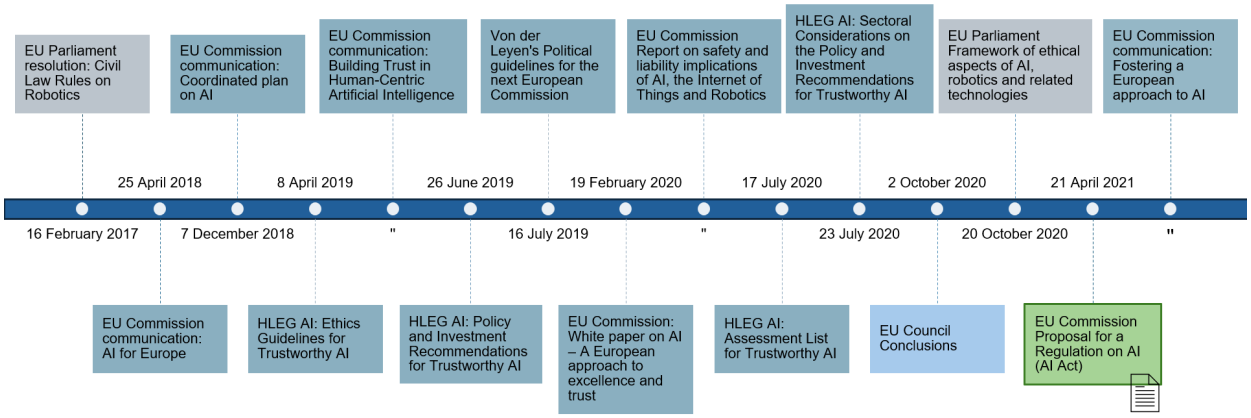

Figure 1: Timeline of policy documents and ethical guidelines published by the European institutions from 2017 to 2021. Representation from the author.

as the AI Act in this respect. After examining its various sources of inspiration, we take a look at the AI Act's provisions in Section 5: its scope, the classification of AI systems and their relative obligations, as well as its proposal for ensuring innovation and monitoring the proper enforcement of requirements. We also map the new governance ecosystem created by the AI Act, as well as the expected deliverables, and indicate important future dates for its entry into force. Finally, in Section 6, we look at the various criticisms that were addressed by the academic literature to the AI Act. While disagreements on the definition of AI, which systems should be prohibited or high-risk and how to implement requirements, are not entirely solved, they are not structural problems. On the contrary, we discuss in Section 7 the risk-based approach chosen by the Commission and how it also attempts to take fundamental rights into account. We show that the distinctive European approach to AI regulation, which blends risks and rights, raises questions about the implementation of the regulation.

2 The growing discussion on AI in Europe

In this section, we present a brief history of the policy documents and ethical guidelines published by the European institutions prior to the AI Act and how they influenced it. To guide the discussion, a timeline is presented in Figure 1.

2.1 AI increasingly became a topic of interest

In 2017, discussions on robotics reached the European institutions with the EU Parliament resolution on Civil Law Rules on robotics (European Parliament, 2017). The text was subsequently strongly criticised by AI and robotics experts, who were particularly concerned about one of the European Parliament's recommendations to the EU Commission, pushing for the creation of a legal status for robots in the long run. For the experts, giving robots legal status was a slippery slope as it would have grant them rights and obligations, blurring the lines between science fiction and reality, and opening the door to liability issues, as any accident caused by the robot would have incurred the liability not of its owner, but of the robot itself (Robotics Openletter, 2017). This idea of giving legal status to a robot or algorithm was latter abandoned and publications started focusing more and more on AI. Since then, AI and digital technologies have been at the top of Europe's agenda. Ursula von der Leyen, then candidate for the presidency of the European Commission, made it one

of her priorities to have *"a Europe fit for the digital age"* (von der Leyen, 2019).

But while the EU Commission has revealed a plethora of different AI policy documents since then, the EU Council only mentions AI in the conclusions of a meeting held in October 2020, where it recognises that the EU needed to be *"a global leader in the development of secure, trustworthy and ethical AI"* (European Council, 2020). It further called on the EU Commission to propose ways of developing research and innovation in the field of AI. The European Parliament, on the other hand, has steered its course on AI through a series of resolutions, generally focusing on sector-specific measures, such as criminal justice or education, and on specific issues raised by AI, such as intellectual property rights or the economic aspects of AI (European Parliament, n.d.a).

In April 2018, the European Commission presented its strategy for AI in a communication entitled "Artificial Intelligence for Europe" (European Commission, 2018a). This strategy was intended as a response to the rapid progress being made by the United States and China in the field of AI, with both countries battling for the lead in the "AI race", with Europe clearly lagging behind (Smuha and Yeung, 2024). The European approach to AI regulation was shaped as a distinct brand, based on European values, positioning itself in contrast to the state-controlled model of China and the permissive model of the US.

The strategy presented relied on four key points moving forward: (1) *"boosting the EU's technological and industrial capacity"* by increasing investments in AI, supporting research, building testing infrastructures and making more data available; (2) *"preparing for socioeconomic changes"* by encouraging diversity and interdisciplinarity and creating an attractive environment for talents in the EU; (3) *"ensuring an appropriate ethical and legal framework"* by drafting AI ethics guidelines and ensuring safety and liability; and (4) *"joining forces"* by engaging both with Member States and AI stakeholders. This communication is at the origin of a number of initiatives that we know today. In particular, it encouraged the drafting of AI ethics guidelines, which would later lead to the creation of the High Level Expert Group on AI (HLEG). The communication also stressed the need to develop standards to *"increase consumer trust"*. As a result, standards now play a major role in the AI Act (Gornet and Maxwell, 2024). Finally, the communication discusses the need to reinterpret the Product Liability Directive in light of technological developments, leading to a proposal for a revision of the framework in 2022, as well as a proposal for a new AI Liability Directive. Both initiatives are still ongoing.

2.2 A European discourse based on trust and respect for fundamental rights

A second communication from the Commission was published in December of the same year, the "Coordinated plan on AI" (European Commission, 2018b), containing actions to be undertaken by the Member States and the Commission. Investment and support resources were specified, with quantified objectives. In particular, a deadline was set for the development of ethical guidelines in March 2019. Both communications emphasised the need for legal rules and ethics guidelines, meant to complement each other and to help protect fundamental rights. The emphasis is on put on "trust", which will later become the cornerstone of all EU deliverables on AI.

Before the legal rules of the AI Act, the ethical framework was the first to be put in place, with the creation of the European Commission's High Level Expert Group on AI (HLEG) in June 2018. The HLEG was tasked with preparing two complementary deliverables: one aimed at AI practitioners, the "Ethics Guidelines for Trustworthy AI" (HLEG, 2019a), and the other addressed to the EU institutions and Member States, the "Policy and Investment Recommendations for Trustworthy AI" (HLEG, 2019b). On the day of release of the Ethics Guidelines, the EU Commission published its third communication on AI: "Building Trust in Human-Centric Artificial Intelligence" (European Commission, 2019). The aim of this last communication was to support the work of the HLEG, by

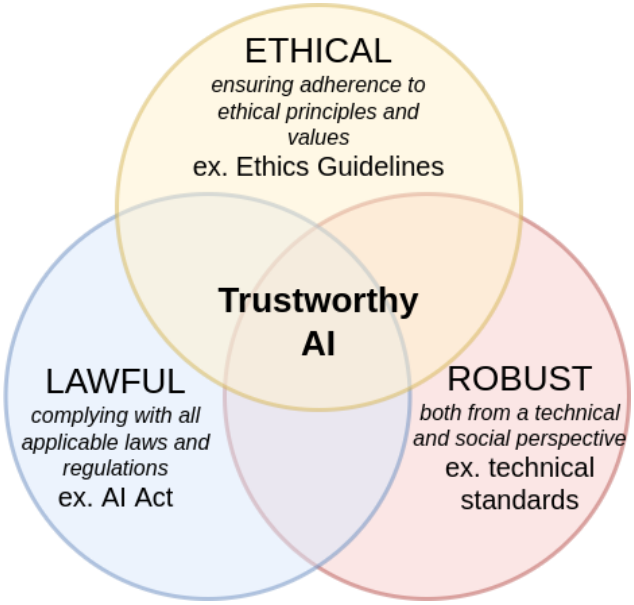

Figure 2: Visual representation of the three pillars of "Trustworthy AI", as intended by the HLEG.

summarising the experts' conclusions and outlining the next steps in updating these guidelines and exporting the EU's expertise in drafting ethical guidelines outside Europe.

Subsequently, the HLEG Ethics Guidelines became one of the most referenced AI ethics documents [1]. In many ways, these guidelines have greatly influenced the field of AI ethics, as well as the discourse of European institutions going forward. The guidelines can therefore be seen as the cornerstone of the European strategy for regulating AI. Notably, they introduced the term "trustworthy AI", which will remain in all of the following European documents, including the AI Act. According to the HLEG guidelines: *"Trustworthy AI has three components, which should be met throughout the system's entire life cycle: 1. it should be lawful, complying with all applicable laws and regulations; 2. it should be ethical, ensuring adherence to ethical principles and values; and 3. it should be robust, both from a technical and social perspective, since, even with good intentions, AI systems can cause unintentional harm"* (HLEG, 2019a).

This enables us to identify the three pillars that the EU has chosen to push forward: the legal sphere with the AI Act, the ethical sphere with ethics guidelines, and the technical sphere with standards and product safety. These pillars, as intended by the Commission, are represented in Figure 2.

After a year's absence, the HLEG was back in force in 2020 with two new publications: the "Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment" (HLEG, 2020a), and "Sectoral Considerations on the Policy and Investment Recommendations" (HLEG, 2020b). Both of these documents were follow ups of their previous documents – the Ethics Guidelines and the Policy and Investment Recommendations. The Sectoral Considerations were primarily geared towards the industry providing AI in three sectors – the IoT sector, the public sector and healthcare, while the ALTAI attempted to overcome one of the biggest challenges of ethical guidelines: operabilty. Indeed, aware that the other charters were merely lists of inoperative principles, one of the HLEG's objectives was to go beyond simply listing ethical principles and to provide guidance on their practical implementation (Smuha, 2019). Alongside the three components for "trustworthy

[1] Based on the study conducted by Gornet et al. (2024).

AI" – lawful, ethical and robust, the Ethics Guidelines therefore introduced an additional list of seven key requirements: (1) human agency and oversight; (2) technical robustness and safety; (3) privacy and data governance; (4) transparency; (5) diversity, non-discrimination and fairness; (6) environmental and societal well-being; and (7) accountability. Each of these requirements was then dissected in the ALTAI into a series of questions addressed to AI practitioners in companies. The ALTAI is designed as a checklist, a tool to support the development of "trustworthy AI" For each requirement, questions are asked on the context in which the system will be deployed, as well as the processes or measures put in place to take the requirement into account. The precision and technicality of the questions asked make the ALTAI a special AI ethics document, halfway between ethical guidelines and technical standards.

The Ethics Guidelines and the ALTAI have notably influenced some of the requirements of the AI Act, such as one of the amendments proposed by the EU Parliament which introduced a Fundamental Rights Impact Assessment (FRIA). At the time the amendment was proposed, the HLEG's "key requirements" were included directly, without change, in the list of criteria for assessing fundamental rights[2].

2.3 The birth of a risk-based approach

The second deliverable of the HLEG, the Policy Recommendations, also had a major influence on the drafting of the AI Act. Indeed, for the first time, the document called EU institutions to adopt *"a risk-based approach to policy making"*, but insisting also on a complementary *"precautionary principle-based approach"* that was later tuned down by the Commission in the proposal for an AI Act (Smuha and Yeung, 2024). Notably, although the HLEG proposed the risk-based approach, it did not advocate the use of the product safety framework for AI (Almada and Petit, 2023). This approach came directly from the Commission when it proposed the AI Act[3].

The Commission continued to emphasise the need for more regulation through two subsequent publications: the "White Paper on AI" (European Commission, 2020b) and a report on the "safety and liability implications of AI" (European Commission, 2020a), both published in February 2020. The report, while acknowledging that safety and liability frameworks already applied to AI products, noted that some risks specific to these technologies were not addressed by current frameworks and that *"additional obligations may be needed"*. The White Paper claimed that the product safety and civil liability legislation was insufficient at the time to deal with AI-related damage, and called for these frameworks to be supplemented. Furthermore, the risk-based approach introduced in the Policy Recommendations and the emphasis on trust from the HLEG guidelines were enhanced by the Commission in the White paper.

The White paper proposed two categories of risks: *"high risk"* and *"not high risk"*. For high-risk AI systems, the White paper lists some provisions that would subsequently become requirements under the AI Act, such as the obligation to use representative datasets, provisions on record keeping, transparency, robustness and accuracy, human oversight, etc. In the White paper, the Commission also recognises the need for prior conformity assessment, which would verify and ensure that the requirements for high-risk applications are complied with. The White paper notably states that: *"the prior conformity assessment could include procedures for testing, inspection or certification. It could include checks of the algorithms and of the data sets used in the development phase"*; and that *"the conformity assessments for high-risk AI applications should be part of the conformity*

[2]This provision is included in (European Parliament, 2023a, amendment 413). In the final text of the AI Act, the list of criteria is not included, but the obligation to carry out a FRIA for high-risk systems is maintained (art. 27 AI Act).

[3]See Section 2.4.

assessment mechanisms that already exist for a large number of products being placed on the EU's internal market". Furthermore, standards are cited in the White paper as a means of facilitating compliance. It is therefore clear that the risk-based structure of the AI Act, as well as parts of its provisions on conformity assessments and standards, is inherited from the White paper[4].

In addition, for AI applications that would not qualify as high-risk, the White paper proposes the use of voluntary labels. Although this is not the solution chosen in the AI Act, it comes close to the use of codes of conducts for "other AI systems"[5]. Finally, the White Paper also stresses the need for a European governance framework, based on national authorities, but also on participation and advice from various stakeholders. The AI Act took these ideas and improved on them, including the creation of new entities which are presented in Section 5.5.

Following the Commission's documents, the European Parliament strengthened the case for more regulation in October 2020 by adopting two Resolutions. In its first Resolution on a "civil liability regime for AI", the Parliament recognised that while a *"complete revision"* of the liability regime was not needed, *"adjustments"* were necessary (European Parliament, 2020a). The Resolution includes a Proposal for a Regulation that has served as inspiration for the Commission's proposal for an AI Liability Directive. In the second Resolution on a "framework of ethical aspects of artificial intelligence, robotics and related technologies", as in the Resolution on a liability regime for AI, the Parliament included a draft proposal for a Regulation (European Parliament, 2020b). Although the framework has been considerably modified by the Commission, the Parliament's proposal is at the origin of what will become the AI Act. However, unlike the Commission's AI Act, and as the title of the Resolution suggests, the Parliament's proposal is strongly based on ethical principles and fundamental rights, such as human oversight, transparency and non-discrimination. Yet, the Parliament seems to mix ethical principles and fundamental rights without making a clear distinction: Article 5 of the proposal is entitled "ethical principles of AI" but deals directly with the fundamental rights enshrined in the EU Charter. In addition to an approach based on ethics and fundamental rights, the Parliament also advocates in the Resolution for a risk-based approach to AI regulation, where compliance would be based on standards. The Resolution states that *"any future regulation should follow a differentiated and future oriented risk-based approach to regulating artificial intelligence, robotics and related technologies, including technology-neutral standards across all sectors, with sector-specific standards where appropriate"*. The Parliament, like the Commission in the White Paper, refers to high-risk AI applications that would be subject to mandatory compliance.

2.4 Towards a mandatory horizontal regulation

The various policy documents published by the three EU institutions – the Commission, the Parliament and the Council – have increasingly moved away from a discourse based on ethics and fundamental rights towards strict regulation based on the risks posed by AI, compliance with which could be assessed by means of technical standards. It was this approach that gave rise to the AI Act. But to explain this transition from ethics, trust and fundamental rights to what is now a product safety regulation, the Commission has released a number of documents alongside the AI Act proposal in April 2021.

First, like all EU regulation, the AI Act was accompanied by an "impact assessment" (European Commission, 2021b), designed to explain the Commission's choice to adopt a certain regulatory approach and why other possible approaches were rejected. Five options were initially considered:

[4]For more information on the risk-based structure of the AI Act, see Section 5.3.
[5]See Section 5.3.

(1) a EU voluntary labelling scheme, (2) an ad-hoc sectoral approach, (3) an horizontal risk-based act, (4) codes of conducts, and (5) an horizontal act for all AI. The impact assessment describes each option and further looks into the advantages and disadvantages of each of them, based on certain themes. In almost all themes, options 3 and 4 are favoured, and the others are found to have too many drawbacks. The labelling scheme option is considered uncertain, with no guarantee that it will be widely adopted, as companies will only agree to undergo a labelling audit if the costs are lower than the benefits. The sectoral approach was also rejected on the grounds that it could lead to inconsistencies in the requirements imposed by sectoral legislation, that regulation would only take place once concerns had been identified, that it would not prevent Member States from adopting their own horizontal regulations, leading to heterogeneous legislation across the EU, and that companies with different AI systems used for different use cases would have to bear multiple compliance costs. As for the last option, which would impose the same strict requirements on all AI systems, although the Commission recognises that it would protect thoroughly the safety and fundamental rights of citizens, it would also expose small businesses to potentially significant compliance costs and create a heavy regulatory burden. On the contrary, the option of a horizontal regulation based on different categories of risk, completed by codes of conduct for low-risk systems, is supposed to *"enhance users' trust"*, and increase legal certainty. However, it should be noted that the core of the impact assessment focuses on the risk-based option, which is the most widely discussed[6]. It therefore seems like the other less developed options were never seriously considered.

In addition to the impact assessment, the proposal on an AI Act also came with an introductory part: the Explanatory memorandum, which offers context, reasons for the proposal and allow us to understand the intent of the Commission[7]. Several of the documents mentioned earlier in this section – such as the HLEG deliverables, the White Paper on AI or the Parliament resolutions – are cited in the explanatory memorandum, as they were major sources of inspiration for the legal text. The Explanatory memorandum gives further arguments to justify the choice of an horizontal approach. We learn that the regulation is intended to be *"comprehensive and future-proof"* with *"flexible mechanisms that enable it to be dynamically adapted as the technology evolves and new concerning situations emerge"*.

These documents justify the European Commission's approach with the AI Act, which is not sector-specific, but risk-specific, with the desire to cover AI in general, but with different requirements depending on the risk category[8]. EU officials have declared that this horizontality guarantees *"functional equivalence"*, whereby obligations and enforcement tools are the same regardless of the sector in which the technology is used (Mazzini and Scalzo, 2023).

2.5 The process of adoption of the AI Act

On 21 April 2021, the European Commission published the first proposal to regulate artificial intelligence in Europe (European Commission, 2021e) which will come to be known as the AI Act. Figure 3 traces the timeline of the text's adoption, from this first proposal to the recent publication of the AI Act in the OJEU.

Along with the AI Act, the Commission launched a consultation period, where stakeholders from various backgrounds[9] were given the opportunity to provide comments on the text. The

[6]The description of all the options runs from page 39 to page 62. The risk-based approach is described on 14 pages, while the other options are described much more briefly, the maximum being the sectoral option, which is described on 5 pages.

[7]The Explanatory memorandum appears on the same document as the Commission proposal and can thus be found here (European Commission, 2021e).

[8]For more information on these different categories, see Section 5.3.

[9]Notably NGOs, academic and research institutions, companies and businesses, and various civil society actors.

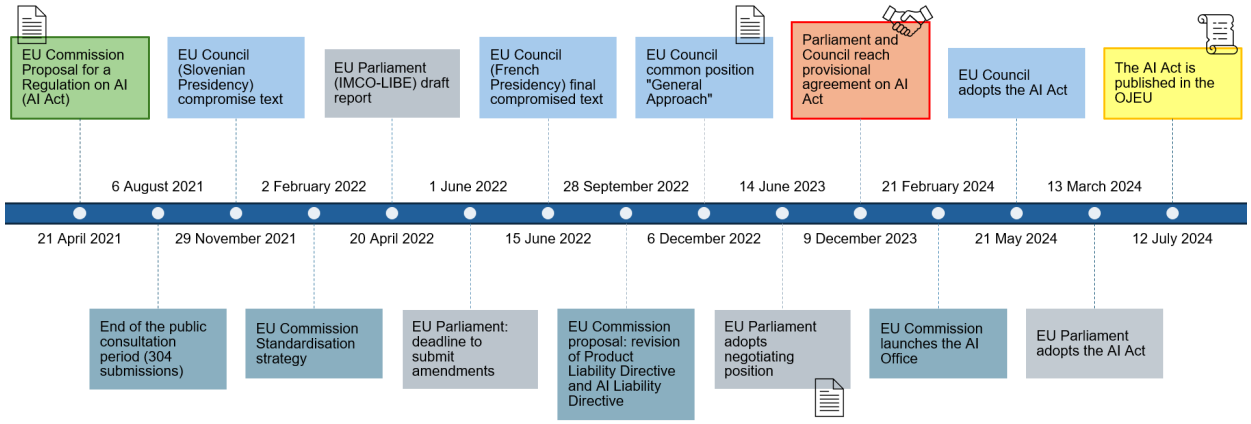

Figure 3: Timeline of the adoption of the AI Act.

Commission received over three hundred submissions[10].

The release of the first draft of the text by the EU Commission was the first step towards the adoption of a mandatory framework for AI in the EU. However, to come into force, the AI Act had still a long way to go. The two other EU institutions, the Council and the Parliament, had to propose amendments to the text. The rotating presidency of the Council meant that one member state would lead efforts to amend the text for six months before another member state took over. The first amendments were therefore proposed at the initiative of the Slovenian presidency and covered only articles 1 to 7, making changes to prohibited and high-risk AI systems in particular (European Council, 2021). The next presidency, led by France, then proposed a large number of changes, notably to article 4, proposing to regulate general purpose AI systems (European Council, 2022b). These contributions were brought together in the French presidency's compromise text (European Council, 2022c). The subsequent Czech presidency continued the process of amending the text, resulting in the General Approach, at the end of 2022 (European Council, 2022a)[11].

The European Parliament operates differently to the Council. The Parliament is made up of different committees, responsible for examining legislative proposals and proposing amendments, which are then submitted in the form of reports to the Parliament who adopts them in plenary session. Two of these committees were chosen to lead the negotiations on the AI Act: the Committee on the Internal Market and Consumer Protection (IMCO) and the Committee on Civil Liberties, Justice and Home Affairs (LIBE) (Ada Lovelace Institute, 2021). Five additional committees[12] adopted their own opinions, with proposed amendments to the AI Act. After an agreement was found between the different committees, the negotiating position was proposed to the Parliament as a whole and adopted on 14 June 2023 (European Parliament, 2023a).

After the adoption of the negotiating positions and proposed amendments, the three EU institutions – the Commission, Council and Parliament, entered a "trilogue" phase to discuss these amendments. After a 3-day "marathon" talk, the EU Commission, Council and Parliament reached a provisional agreement on the text on 9 December 2023 (European Parliament, 2023b). The text then underwent a series of minor textual improvements and was approved by the Parliament on 13 March 2024 (European Parliament, 2024) and by the Council on 21 May 2024 (European Council,

[10] All submissions for the AI Act can be found on the Commission's website: (European Commission, 2021a).

[11] A more precise timeline is given by (Future of Life Institute, n.d.).

[12] Namely, the Committee on Legal Affairs (JURI), the Committee on Industry, Research and Energy (ITRE), the Committee on Culture and Education (CULT), the Committee on the Environment, Public Health and Food Safety (ENVI) and the Committee on Transport and Tourism (TRAN)

2024). The text of the AI Act was subsequently published in the OJEU on 12 July 2024 (European Parliament and Council, 2024).

2.6 A continuity of the trustworthiness discourse after the AI Act

Communications and policy documents on AI did not stop with the publication of the AI Act. A wide range of documents continue to be published and EU institutions continue to launched projects. These initiatives explain the European approach to AI regulation, picking up the main key elements and providing new details on the way forward.

First, the Commission released alongside the AI Act a Communication on "Fostering a European approach to AI" to summarise the main elements present is the legal text (European Commission, 2021c). The AI Act is notably said to *"combine greater safety and fundamental rights protection while supporting innovation, enabling trust without preventing innovation"*.

Other documents provide guidance for EU AI policy. For example, the revision of the "Coordinated Plan on AI" sets out the next steps of the EU's strategy for AI (European Commission, 2021d). It is described by the Commission as the *"next step in creating EU global leadership in trustworthy AI"*. The coordinated plan sets several goals: to enable AI development and uptake, foster research excellence, promote the EU vision of *"AI for people"* and as a *"force for good in society"*, and strengthen leadership in key sectors, such as environment, robotics, health, public sector, law enforcement, mobility and agriculture. In particular, since the first version of the coordinated plan in 2018 (European Commission, 2018b), the Commission is committed to opening a small number of *"specialised large-scale reference sites"* across Europe, equipped with technology infrastructures and specific expertise: the AI Testing and Experimentation Facilities (TEFs) (European Commission, n.d.e). Since then, a few collaboration projects have been launched.

In addition, the Commission also published in January 2024 a Communication on "boosting startups and innovation in trustworthy AI" (European Commission, 2024c). It describes new initiatives to support AI startups and SMEs, including the launch of *"AI Factories"*, i.e. computing facilities, resources and services to attract AI *"talents"*. Other initiatives include a number of research and investment programmes.

The work to successfully implement the AI Act will also continue with the AI Office, a new executive organ of the European Commission created by the AI Act[13]. Initiatives lead by the AI Office include the "AI Pact", a voluntary framework towards the industry to anticipate and prepare for future compliance with the AI Act (European Commission, n.d.b). A first call of interest was launched in November 2023. The AI Office then released the AI Pact commitments in September 2024, inviting participating companies to endorse this non-binding framework and report on their progress later. By signing up this Pact, the companies notably pledge to adopt an AI governing strategy, to identify their high-risk systems[14], and to promote AI literacy among staff.

3 The legal landscape of the AI Act

3.1 Digital constitutionalism in Europe

The AI Act will not apply in a vacuum. These past few years, the European Union has produced a proliferation of texts designed to regulate both new technological products and industrial players

[13] See Section 5.5.

[14] High-risk systems are a specific category of AI systems under the AI Act. For more information on the different categories, see Section 5.3.

in the digital age. Some of these texts predate the AI Act, the best known being the GDPR (European Parliament and Council, 2016b) for the protection of personal data and the Data Services Act (DSA) (European Parliament and Council, 2022) and Digital Markets Act (DMA) (European Parliament and Council, 2022a) for the regulation of online platforms. Other texts are still in the making, such as the AI Liability Directive (European Parliament and Council, 2022), or the revision of the e-Privacy Directive (European Parliament and Council, 2017a). In total, there are dozens of texts which, if adopted, will regulate digital technologies in Europe, and the number of legislative proposals is likely to increase still further. [15] The objective of these texts is, among other things, to protect the fundamental rights of EU citizens, which is why scholars have been referring to this trend as *"digital constitutionalism"* (De Gregorio, 2021), i.e. an *"ideology that aims to establish [...] a normative framework for the protection of fundamental rights and the balancing of powers in the digital environment"* (Celeste, 2019). The AI Act is therefore part of this European approach to new technology regulation and will work alongside these other texts in the European legal landscape – some of which are quoted directly in the AI Act.

To navigate this legal landscape, we created a diagram showing the texts which are likely to intersect with the AI Act. This representation is illustrated in Appendix A. It shows all the treaties, directives and regulations cited in the AI Act, organised into large families corresponding to thematic spheres. The diagram is not intended to be an exhaustive list, but rather to give an idea of the multitude of texts involved. In the next section, we take a look at a selection of important texts.

3.2 Relative treaties, directives and regulations

The EU treaties

The AI Act, like all European legislation, is based on the EU Treaties. The two core treaties of the EU, the Treaty on European Union (TEU) (European Commission, 2012b) and the Treaty on the Functioning of the European Union (TFEU) (European Commission, 2012c), define how the EU operates. In accordance with the ordinary legislative procedure established by the Treaty of Lisbon, all new EU legislation must have a legal basis from one of the articles of these treaties[16]. For the AI Act, the Commission motivated the proposal on the basis of data protection (art. 16 TFEU), and functioning of the internal market (art. 114 TFEU). The Charter of Fundamental Rights of the European Union (CFREU) (European Commission, 2012a) is an additional text to be taken into consideration, especially as AI systems represent a danger for individuals and society as a whole.

Personal data protection

Europe also lead the way in personal data protection regulation with the GDPR (European Parliament and Council, 2016b) in 2016. The text lays down rights for data subjects and obligations from data controllers and data processors. Data must notably be processed in a transparent and secure manner, and for limited purposes[17]. In the context of law enforcement, the GDPR does not apply, but a second text, usually generally to as the "law enforcement directive" (European Parliament and Council, 2016a) takes over. Finally, a third text lays down obligations for data processing by the European institutions (European Parliament and Council, 2018).

[15] See a list of future European texts at (Zenner et al., 2024).
[16] See a list of these legal bases in (European Parliament, n.d.b).
[17] For more information on the GDPR and how it served as a source of inspiration for the AI Act, ses Section 4.

Data sharing

But not all data is personal and is covered by the GDPR. To facilitate the sharing in Europe of industrial data and notably data from the Internet of Things (IoT), the European Union adopted in 2022 and 2023 the Data Governance Act (European Parliament and Council, 2022b) and the Data Act (European Parliament and Council, 2023b). In particular, these laws allow data from connected devices to be made accessible, provide for the use of company data by public bodies in exceptional circumstances, abolish fees for changing service providers and offer a number of guarantees against illicit access to data by the governments of third countries.

Digital platforms

After the GDPR and data protection, the European institutions have been tackling the issue of the major digital platforms, with the publication of the DMA (European Parliament and Council, 2022) and DSA (European Parliament and Council, 2022a) in 2022. But the two texts serve very different objectives. The goal of the DMA is to complement competition law to prevent the monopoly of very large platforms – social networks, web browsers, etc. In particular, it aims to facilitate unsubscribing and interoperability with competitors, prohibits self-referencing, forces platforms to inform users on future fusions, and requires consent for the re-use of personal data for targeted advertising purposes. On the other hand, the DSA is focused on internet service providers, cloud services and online platforms. It aims to harmonise regulations on illegal content and products, such as hate speech, child pornography, terrorism, disinformation, drugs, counterfeit goods, and so on. It requires platforms to have a tool for reporting content and handling complaints. It provides for a right to explanation of algorithms, prohibits advertising targeted at minors and provides for risk analysis and annual audits and (limited) access to the interface.

Product safety

The AI Act is a product safety regulation, part of the New Legislative Framework (NLF). AI products are therefore *de facto* covered by the General Product Regulation and, for systems which do not fall into specific categories under the AI Act[18], they will at least be covered by the requirements of their sector-specific regulation. At the time of writing[19], there are 27 directives and regulations aligned with, or based on the NLF[20]. One key inspiration for the AI Act was the Medical Device Regulation (European Parliament and Council, 2017b), which is broadly seen by Europe as the success story of product safety regulation (Mazzini and Scalzo, 2023). Other example of product safety regulations include for instance the Toy Safety Directive (European Parliament and Council, 2009), the Machinery Regulation (European Parliament and Council, 2023a) and the Radio Equipment Directive (European Parliament and Council, 2014).

Product liability

The AI Act is an *ex ante* regulation: it sets out the requirements that AI systems must meet *before* being placed on the market, but it does not cover *ex post* liability rules. Pending the specific rules set out in the AI Liability Directive (European Parliament and Council, 2022), the AI Act will for the time being follow the General Product Liability Regulation (European Council, 1985).

[18] See Section 5.3.
[19] November 2024.
[20] A list of these directives and regulations can be found at (European Commission, n.d.d).

4 The GDPR: the predecessor for digital rights

4.1 A risk-based approach initiated by the GDPR

Among the texts of the European legal ecosystem, one stands out: the GDPR. As the first legal text adopted in Europe to tackle digital technologies directly, the GDPR has strongly influenced AI Act in its spirit and structure, and paved the way for future digital laws.

And indeed, the AI Act takes inspiration from the GDPR on a number of points. The GDPR's main objective is to protect citizens' rights to privacy and data protection. But while protecting these rights, the GDPR has also introduced the beginning of a risk-based approach that will be taken up and enhanced in the AI Act.

In the GDPR, the data controller[21] must carry out a Data Protection Impact Assessments (DPIAs) for high-risk processing operations (art. 35.1 GDPR). This includes profiling, large scale processing and systematic monitoring (art. 35.3 GDPR). The assessment must contain a description of the operations, an assessment of the necessity and proportionality of operations, an assessment of the risks to rights and freedoms of data subjects, and the measures envisaged to address these risks (art. 35.7 GDPR). When the DPIA shows that the processing presents a high-risk in the absence of mitigating measures, the supervisory authority must be consulted (art. 36.1 GDPR). The GDPR's DPIAs have notably influenced the FRIAs of the AI Act.

In addition to its novel risk-based approach, the GDPR has initiated a shift from a system of static prior formalities to a system of dynamic global compliance. Businesses do not need prior authorisation from supervising authorities to process personal data as before, but instead must be able to demonstrate at any time that they are complying with the principles of the Regulation. This is best shown in Article 24 on the responsability of the controller where he or she must *"ensure and be able to demonstrate that processing is performed in accordance with [the] Regulation"* by means of technical and organisational measures. Recital 74 goes even further, asserting that this obligation to demonstrate compliance also applies to the effectiveness of the measures. The notion of compliance in the GDPR is thus very much intertwined with the notion of risk[22], the lack of compliance creating more risks to the data subjects' right to privacy.

4.2 The importance of the state of the art in the GDPR

The state of the art plays an important role in the GDPR, a role that will be further strengthened by the AI Act by relying on harmonised standards. Indeed, the GDPR makes trade conditional to the fulfillment of certain obligations by the controller, which are directly defined in the text. For instance, under the principle of integrity and confidentiality, the controller must ensure *"appropriate security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage"* (art. 5.1(f) GDPR). However, the means of achieving this goal in practice are left to the controller, who must take *"appropriate technical or organisational measures"*, but none are cited in the text. This security obligation is further strengthened by Article 32 on the security of processing, which stipulates that *"the controller and the processor*[23] *shall implement appropriate technical and organisational measures to ensure a level of security appropriate to the risk"*. Unlike Article 5, however, Article 32 gives broad examples of how these measures can be carried out, such as the pseudonymisation and encryption of personal

[21] The controller is the natural or legal person who determines the purposes and means of the processing of personal data (art. 4(7) GDPR).

[22] Gellert (2018) speaks of a *"compliance risk"*.

[23] The processor is the natural or legal person who processes personal data on behalf of the controller (art. 4(8) GDPR).

data. Yet, these measures remain very general and no concrete information is given on the technical aspects of their implementation. It is up to the controller and processor to choose what measures to put into place to ensure and demonstrate compliance with the Regulation (art. 24.1 and 28.1 GDPR).

One key requirement introduced by the GDPR is that privacy and data protection need to be integrated in the technology when it is created, which is referred to as data protection "by design" and without human intervention, known as data protection "by default" (art. 25 GDPR). This approach by design is also included in the AI Act, particularly in the requirements for high-risk systems: mitigating risk management requires *"adequate design and development"* (art. 9.4 GDPR) and data management includes *"relevant design choices"* (art. 10.2(a) GDPR). Furthermore, high-risk systems must be *"designed and developed"* to enable the recording of events (art. 12.1 GDPR), the transparency of the system and the interpretation of outputs (art. 13.1 GDPR), an effective oversight (art. 14.1 GDPR) and an appropriate level of accuracy, robustness and cybersecurity (art. 15.1 GDPR). This approach by design notably requires consideration of the state of the art, as noted by Article 25.1. For instance, the processing security measures required by Article 32 must be taken into account the state of the art.

4.3 Standards for the GDPR

However, this does not mean that there is no technical standards for the GDPR. The most well know standard to tackle issues close to that of the GDPR is ISO/IEC 27701 (2019), based on two information security standards: ISO/IEC 27001 (2022a) on information security management systems and ISO/IEC 27002 (2022b) on security measures. However, while applying these standards is a way for companies to show that they are implementing good practices and can be seen as a first step towards compliance with GDPR requirements (Lopes et al., 2019), it remains insufficient to demonstrate full compliance with the European law. In particular, the French data protection authority, the CNIL, has stated that these standards are not GDPR specific and cannot be considered as a valid certification scheme for the GDPR, although they do represent the state of the art (CNIL, 2020). Indeed, ISO standards are international standards, far removed from European concerns about privacy protection. The first version of ISO/IEC 27001 for instance was published in 2005 (ISO/IEC, 2005), well before the GDPR was adopted in Europe.

To bridge the gap between insufficient international standards and GDPR requirements, the European Commission mandated the European standardisation organisations to prepare standards for the GDPR in a standardisation request on "Privacy and personal data protection management" delivered in 2015[24]. A few standards have been developed following that request[25]. These include standard EN 17529:2022 on "Data protection and privacy by design and by default" developed by CEN-CENELEC Joint Technical Committee 13 on "Cybersecurity and Data Protection". However, this standard, although adopted at European level, is not expected to be cited in the OJEU for the GDPR[26].

While the GDPR is not based on compliance with standards, unlike the AI Act, it is nevertheless one of the first times that the Commission has requested standards for a European law in the digital field. Additionally, the EU considers privacy and data protection to be a fundamental right, enshrined in Article 8 of the Charter of Fundamental Rights (CFREU) (European Commission, 2012a). The standardisation request on privacy and personal data protection management was

[24]The reference to such request can be found in (European Commission, 2016). It is referred to as standardisation request M/530. However, we were unable to find the full text of the request.

[25]See a list of these standards at (ITEH Standards, n.d.).

[26]See in particular the standard page on the CEN-CENELEC website: (CEN, n.d.).

subsequently based on this same article (Kamara, 2017). The CEN-CENELEC JTC 13 standards are therefore the first attempt to develop standards relating to fundamental rights, something the AI Act aims to achieve on a larger scale (Gornet and Maxwell, 2024).

4.4 Voluntary certification in the GDPR

Even if some technical standards exist for the GDPR, unlike the AI Act, the GDPR does not provide for the use of CE marking, as this mechanism is generally reserved for products covered by the NLF. However, the GDPR uses voluntary certification, based on co-regulatory tools: codes of conduct and certification mechanisms, which play an important role in making controllers and processors accountable. These replace the traditional privacy seals[27], but are not published in the OJEU and do not benefit from the advantages of harmonised standards[28] and CE marking schemes.

Associations and bodies representing data controllers or processors in a given sector of activity can draw up codes of conduct to apply the GDPR rules to their sector or adapt them to the specific needs of enterprises (art. 40.1 GDPR). Compliance with a code of conduct is voluntary, yet strongly encouraged, and is based on a self-certification mechanism with subsequent monitoring (art. 41.1 GDPR). Other standardised mechanisms include binding corporate rules (art. 47 GDPR) which allows companies to transfer personal data in and out of Europe, but that are mandatory once signed.

Furthermore, Article 42.1 provides for the *"establishment of data protection certification mechanisms and of data protection seals and marks"*. The purpose of these tools is to enable controllers and processors to demonstrate that their personal data processing complies with the Regulation. They can also be used to justify that a company that is not subject to the obligations of the GDPR complies with its principles and presents appropriate guarantees in the event of data being transferred outside the Union (art. 42.2 GDPR). This is a voluntary process requiring an *a priori* assessment by an accredited certification body[29], a supervisory authority[30] or the European Data Protection Board (EDPB)[31]. If the EDPB considers criteria of a specific certification scheme consistent with the GDPR, this will result in a common certification called the "European Data Protection Seal" (art. 42.5 GDPR).

In this context, the EDPB adopted in October 2022 an Opinion on the Europrivacy (n.d.a) criteria for certification (EDPB, 2022b). This marks the approval of the very first European Data Protection Seal (EDPB, 2022a). Companies can now evaluate their data processing against the Europrivacy criteria and show their compliance to the GDPR. With this status, Europrivacy certificates will be officially recognised in all EU countries. The Europrivacy certification covers the ISO/IEC 27001 standard (Europrivacy, n.d.b) which is already largely used, making the international criteria one of the components of this official European certification.

Certification thus plays a key role in the GDPR for the accountability of stakeholders, as adherence to approved codes of conduct or approved certification mechanisms may be used as an element

[27] A privacy seal is *"a certification mark or a guarantee issued by a certifying entity verifying an organisation's adherence to certain specified privacy standards that aim to promote consumer trust and confidence in e-commerce"* (Rodrigues et al., 2013).

[28] Harmonised standards are European standards that can be published in the OJEU and be granted certain legal properties (Gornet and Maxwell, 2024).

[29] Certification bodies issue and renew certification. They are accredited by either a supervisory authority or a national accreditation body (art. 43.1 GDPR). To be accredited, a certification body must notably demonstrate its *"independence and expertise in relation to the subject-matter"* (art. 43.2(a) GDPR).

[30] A supervisory authority is an independent public authority established by a Member State (art. 4(21) GDPR) and who is responsible for monitoring the application of the Regulation (art. 51.1 GDPR).

[31] The EDPB is an independent Union body (art. 68.1 GDPR) responsible for ensuring the consistent application of the Regulation (art. 70.1 GDPR).

by which to demonstrate compliance with the obligations of the controller and processor (art. 24.3 and 28.5 GDPR). Moreover, multiple certification models co-exist in the GDPR and the text does not prohibit the establishment of certification schemes outside of Article 42 regime (Lachaud, 2020). However, some have argued that the certification process under the GDPR could already be seen as a new regulatory instrument (Lachaud, 2018), a form of self-regulation. Yet it scope remains limited and it should be seen as a stepping stone towards the extend of certification provided for in the AI Act.

It should be noted that contrary to the AI Act, certification in the GDPR is issued to a data controller or processor (art. 42.7 GDPR), not to an infrastructure or product. Another difference is the body responsible for the certification assessment: while in the AI Act the conformity assessment is issued internally in some cases, in the GDPR, in the absence of hENs, the assessment will always be carried out by an external certification body. The main advantage of the certification mechanisms provided for by the GDPR is that they give a competitive advantage to companies that comply with them (Grafenstein, 2022). Despite these differences, the use of voluntary certification under the GDPR has therefore paved the way for prescriptive certification in the form of CE marking in the AI Act.

5 Navigating the requirements of the AI Act

5.1 The scope of the AI Act

The AI Act is an EU Regulation, which means it is directly applicable by Member States without the need to transpose it into national laws. The AI Act applies to various stakeholders across the EU and the European Economic Area (EEA)[32], including providers and deployers of AI systems. The provider is the entity responsible for developing the AI system (art. 3(3) AI Act), while the deployer is the entity who uses the AI system (art. 3(4)). AI system providers and deployers must ensure that their AI systems comply with the various requirements of the AI Act[33], depending on the category to which their system belongs[34]. Other stakeholders, such as distributors and importers, also have obligations when the AI system presents a high risk[35]. For instance, they must verify the various stages of the conformity assessment procedure (art. 23 to 26).

The AI Act sets out two main purposes and two ways of achieving them. Indeed, the AI Act seeks both to *"improve the functioning of the internal market"* and to *"promote the uptake of human-centric and trustworthy AI"*. These objectives can be achieved by two main means: *"ensuring a high-level of protection of health, safety, fundamental rights"* and *"supporting innovation"* (art. 1.1). These two objectives are very distinct, and each of the provisions in the AI Act attempts to address one or the other. In particular, the text adopts an approach to product safety as a means of improving the internal market, while adapting it to the protection of fundamental rights. The AI Act therefore conveys a dual discourse: improving the internal market and protecting fundamental rights

To achieve these two objectives, the AI Act lays down various rules and requirements which apply in different contexts. However, for a system to be primarily covered by the AI Act, it must

[32]In Europe, the twenty-seven countries of the European Union and the four countries of the European Free Trade Association (EFTA), namely Iceland, Liechtenstein, and Norway, are bound by the same rules governing the internal market and enabling the free movement of persons, goods, services, and capital within what is known as the "European single market".

[33]In particular with regard to articles 16, 50, 53 and 55 of the IA Act for the provider, and articles 26 and 50 for the deployer.

[34]See next paragraph for the different categories of AI systems.

[35]See the risk classification in Section 5.3.

correspond to the material and territorial scope of the text. First, the definition of an AI system under the AI Act covers most approaches known as "AI" in computer science. Second, the AI Act applies specifically where a provider places an AI system on the EU market, whether or not it is located in the EU. It also applies when an AI system is used by a deployer whose registered office is located in the EU.

Stakeholders who fail to comply with the provisions of the AI Act may be subject to financial penalties. This fine can be as high as €35,000,000 or 7% of their total worldwide annual turnover for putting on the market a prohibited system. Other fines include, for example, €15,000,000 or 3% of the annual turnover for not complying with their obligations, and €7,500,000 or 1% of their worldwide annual if they fail to cooperate with national authorities.

5.2 Defining AI

AI systems

The AI Act gives a definition of an *"AI system"* (art 3(1) AI Act): *"A machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".*

The definition uses fairly technical terms such as *"inputs"* and *"outputs"*, but what seems to separate a simple algorithmic system from an AI system under the AI Act is the notion of *"autonomy"*[36]. Although we have no definition of this term in law, in a general sense it can mean: *"the quality or state of being self-governing"* (Merriam-Webster Dictionary, n.d.).

With this definition alone, it is not clear whether symbolic AI models[37] are included or just machine learning models. Indeed, even if the objectives can be *"implicit"* or *"explicit"*, the definition insists on the need for inference. This is further exacerbated by Recital 12, which states that the definition of AI systems *"should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations"* and that *"a key characteristic of AI systems is their capability to infer"*. However, the same recital goes on to state that: *"The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved"*. Therefore, both Symbolic AI and ML methods should be covered by the AI Act[38].

Systems or models?

The AI Act makes a point of regulating specifically AI *"systems"*, not AI *"models"*, with the exception of General Purpose AI models[39]. The AI Act explains the difference between AI systems and AI models, as Recital 97 states that: *"Although AI models are essential components of AI systems, they do not constitute AI systems on their own. AI models require the addition of further components, such as for example a user interface, to become AI systems. AI models are typically integrated into and form part of AI systems"*. Recital 101 further emphasises that models can be integrated into products, but do not constitute products in themselves. This is why models are only mentioned when training is involved (Recitals 67 and 76). *"Systems"* are the end products, while

[36] This definition has evolved greatly since the first proposal of the Commission. See Section 6.1.

[37] Which function on explicit rules defined by human agents.

[38] However, this conclusion will have to be verified in the future, when we have further case law on which to rely.

[39] See next paragraph.

"models" enable the systems to function. The main reason for this distinction is that the AI Act aims to regulate only products that are put on the market, so they must be final end products and not components of such products.

General Purpose AI and Generative AI

The AI Act makes a distinction between traditional AI systems and *"General Purpose AI"* (GPAI) systems. GPAI systems are defined as: *"an AI system [...] which has the capability to serve a variety of purposes"*.

GPAI systems are notably based on GPAI models, defined in Article 3(63) as: *"An AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications"*.

The definition insists on the importance of the wide variety of tasks such a system can solve and the wide range of applications to which it can be applied. We can also note a desire to describe in more detail how GPAI models can work, specifying that this applies to systems working on large amounts of data and using self-supervised learning[40].

In particular, the notion of GPAI in the AI Act corresponds to the computer science concept of *"foundation model"*[41], a name which was briefly used some amendments to the AI Act[42], before being removed from the final version.

Finally, while GenAI does not have a specific definition, *"AI systems generating synthetic audio, image, video or text content"* are cited in the text (art. 50 AI Act), and should not be confused with GPAI systems, a category of systems to which they may or may not belong. Indeed, while most GenAI systems today are capable of processing different types of data, we could imagine a system which only generates a certain type of data.

5.3 The risk based approach of the AI Act

The AI Act is a risk-based regulation, meaning AI systems are classified into certain categories of risks and for each category, certain requirements apply. When the Commission released its first proposal, its representation of the risk hierarchy followed a pyramid shape with unacceptable risk systems at the top of the pyramid, followed by high risk, limited risk and, finally, minimal risk systems at the bottom, as shown in Figure 4.

As amendments have been made, the structure of these categories has evolved. Although the final version of the regulation retains its risk-based structure, the risk categories have changed and certain types of risk can be accumulated, making the pyramid representation obsolete. We propose an alternative representation in Figure 5.

The two main categories of risk remain: unacceptable risk AI systems, and high-risk AI systems, now respectively at the top of our representation.

[40] For a technical definition of those terms, see Section **??**.

[41] To understand how foundation models work, see Section **??**.

[42] The term *"foundation model"* was introduced by the European Parliament during the negotiations on the IA Act and appears in various amendments to the text. A definition is notably given in amendment 168 (European Parliament, 2023a).

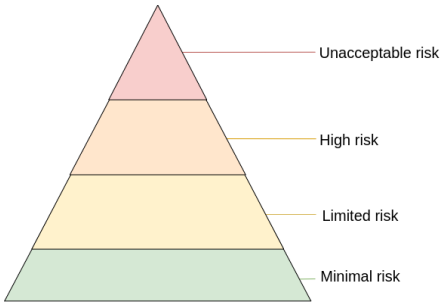

Figure 4: Pyramid of risks as initially intended by the Commission in the first draft of the AI Act. This representation, although still used by the European institutions, is now depreciated in the final version of the text. Image adapted from (European Commission, n.d.a).

Unacceptable risk

The use of AI systems which present an unacceptable risk is totally prohibited by the AI Act. For these systems, the EU has opted for a precautionary approach (Almada and Petit, 2023). Unacceptable risk systems include – but are not limited to – social scoring[43], predictive policing[44], emotion recognition in the workplace or the education system[45], biometric classification[46] and biometric identification, including facial recognition, under certain conditions[47] (art. 5.1).

High risk

There are two ways to fall into the high-risk category: either (i) the product using AI is already covered by EU harmonised legislation; or (ii) the domain of application of the AI system must be listed in Annex III (art. 6). The EU harmonised legislation for (i) is listed in Annex I. It contains the twelve NLF regulations, such as those on machinery, toys, lifts, radio equipment and medical devices, as well as other legislation covering, for instance, certain motor vehicles or aircraft. The AI system must be either the safety component of a product covered by one of the regulations, or itself a product covered by the regulation. In addition, it must undergo a conformity assessment by a third party in accordance with this regulation (art. 6.1). In addition, for (ii), systems covered by Annex III, called *"stand-alone AI systems"* (rec. 52) include biometric systems, AI systems used for critical infrastructure, education, employment, essential public services, law enforcement, migration and justice. There are, however, exceptions whereby systems listed in Annex III may not be considered high risk, for instance if they are intended to perform a *"narrow procedural task"* or to simply improve the result of a human activity (art. 6.3). All AI systems listed in Annex III will be registered in an EU database (art. 71.1). The high-risk category should cover approximately 5 to 15% of all AI systems in the EU, according to the European Commission (2021b). However, other studies show that this figure could actually be much higher (appliedAI, 2023).

High risk AI systems have to comply with a list of essential requirements that can be found through Articles 9 to 18. Under these requirements, the provider should: establish a risk man-

[43]Systems which evaluate the social behavior of individuals.
[44]Systems which predict the risk of a person to commit a criminal offence.
[45]Unless for medical or safety reasons.
[46]Systems designed to deduce protected characteristics of individuals, such as race, political opinion, religious belief, sexual orientation and so on, by using their biometric data.
[47]When it is used in real-time and remotely, in a public space, for law enforcement purposes, and when it does not fall under a list of exceptions such as the targeted search for victims of human trafficking or perpetrators of serious criminal offences, or the prevention of terrorist attacks.

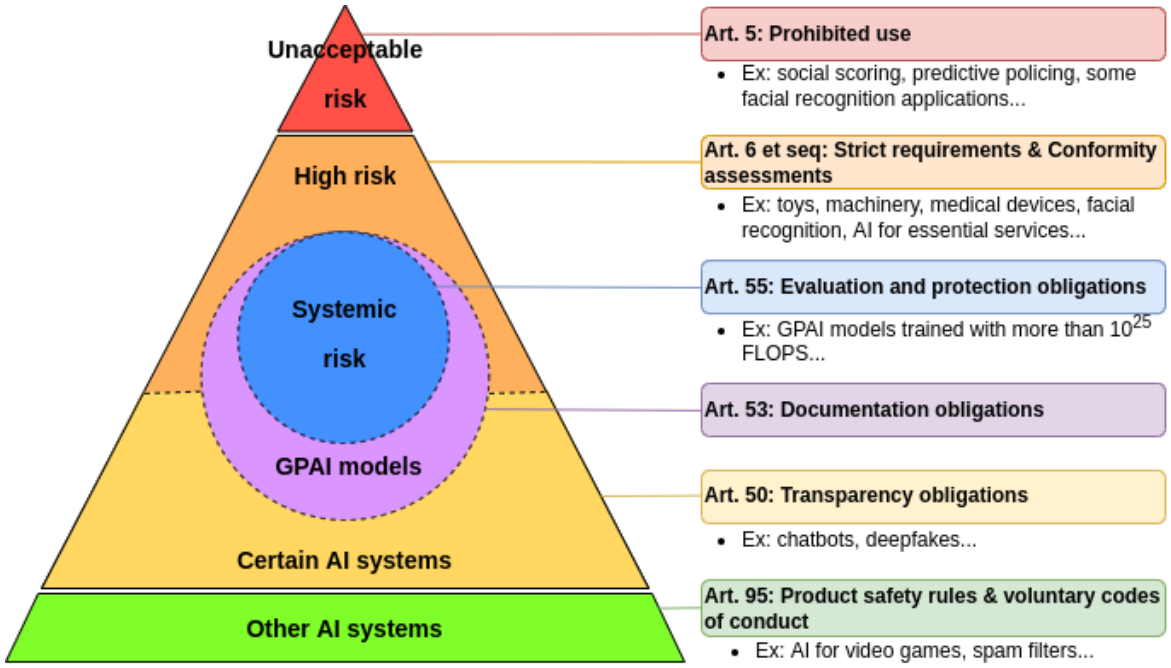

Figure 5: Visual representation of the categories of AI systems in the AI Act (by the author). Dotted lines represent categories which can be accumulated.

agement system and a quality management system (art. 9 and 17), use quality data (art. 10), draw up technical documentation and make it available to national authorities (art. 11 and 18), record events in logs and keep the logs for an appropriate period of time (art. 12 and 19), ensure a level of transparency which enables output interpretation (art. 13), design systems to be overseen by humans (art. 14), ensure accuracy, robustness and cybersecurity of the system (art. 15). In addition, public entities or private bodies providing public services must carry out an assessment of the AI system's impact on fundamental rights (art. 27).

To be distributed on the EU market, high risk AI system must undergo a conformity assessment procedure and receive a European Conformity (CE) mark which shows compliance with the regulation. This procedure may be carried out by a third party or be a self-assessment carried out directly by the company, depending on the application of the AI system. To demonstrate compliance, providers will rely in particular on the state of the art and on harmonised technical standards.

Certain AI systems

The limited risk category first proposed by the European Commission is replaced by the "certain AI systems" category in the final version of the text. This category is cumulative with the category of high-risk systems. This change to the pyramid structure is due mainly to the inclusion in the text of General Purpose AI (GPAI) systems, capable of performing a wide range of tasks[48], and which, depending on their application, may or may not fall into the high-risk category.

The "certain AI systems" category thus includes GenAI systems as a whole[49] – including GPAI

[48]GPAI *systems* are based on GPAI *models*.

[49]The term GenAI is not directly used by the AI Act, which refers to AI systems *"generating synthetic audio, image, video or text content"* (art. 50.2).

systems; emotion recognition or biometric categorisation systems whose application does not fall under the high-risk classification; and AI systems in general which are *"intended to interact with natural persons"*. These systems are subject to transparency obligations. Providers are required to inform individuals when they interact with an AI system and deployers of emotion recognition or biometric categorisation systems must inform individuals when they are subject to the operation of these systems. For GenAI, providers must ensure that the results of their GenAI system can be detected as artificially generated and GenAI deployers must disclose that content has been artificially generated.

GPAI models

GPAI models now have their own category which is also cumulative with others. Indeed, GPAI models can be integrated into AI system and thus fall into, at least, the "certain AI systems" category and, for certain applications, may also fall into the "high risk" category. GPAI models, regardless of the other risk categories in which they may be classified – certain AI systems and/or high-risk – have additional obligations. Providers must: (i) draw up technical documentation, including on the training and testing processes and results, and make it available upon request to the AI Office; (ii) provide information and documentation to providers who intend to integrate the GPAI model into their AI system; (iii) put in place a policy on copyright; and (iv) make publicly available a detailed summary about the content used for training. To show compliance with these requirements, providers can rely on codes of practices approved by the Commission, or on harmonised standards if they exist (art. 53.4).

Systemic risk

In addition to the requirements specific to GPAI models and the requirements linked to other risk categories – certain AI systems and/or high risk – GPAI models which present "systematic risks", due to their scale and the importance of their potential impact, are subject to additional requirements. This is notably the case of models which use an amount of computation of more than 10^{25} floating point operations (FLOP)[50] for training. However, if any models above this threshold is automatically considered with systemic risk, it is a necessary condition to fall in this category. Indeed, high impact capabilities can also be evaluated on the basis of *"appropriate technical tools and methodologies"*. The Commission might decide, following a alert from the scientific panel that a GPAI model presents a systemic risk. This can be assessed using the criteria in Annex XIII, such as the number of model parameters, the quality or size of the dataset and the amount of computation. The type of input and output modality should be taken into account and compared with the state of the art in the field[51]. The Commission could also examine the impact on the internal market and the number of end-users.

Additional requirements for GPAI models with systemic risk fall on the provider, who must evaluate the model using standardised protocols, including adversarial testing, mitigate risk, keep track of and report serious incidents to the AI Office, and ensure cybersecurity protection. To show compliance with these requirements, as with the other GPAI models, providers of GPAI models with systemic risk can rely on codes of practice or harmonised standards (art. 55.2).

Ultimately, some AI systems will fall into up to four categories, each with its own specific requirements. This is the case for GPAI models (GPAI models and certain AI systems category),

[50] A FLOP is a simple mathematical operation, such as addition or division, performed with floating-point numbers, which are approximations of decimal numbers.

[51] The state of the art, for example in FLOP, can vary between text and image generation.

with high capabilities (systemic risk category), which are used for high risk applications (high risk category). It is worth noting, however, the difference between the GPAI model and the AI system in which it is integrated downstream. The GPAI model and the downstream high-risk system may have different providers, each with their own obligations.

Other AI systems

There is no official name for the category of AI systems which are neither unacceptable risk, high-risk, nor GPAI systems. Indeed, other AI systems are not subject to any specific requirements under the AI Act. However, they are mentioned in the text, which notably encourages providers and developers to draw up of voluntary code of conduct (art. 95). In addition, it should be noted that all AI systems fall under the General Product Safety Regulation (European Parliament and Council, 2023c). Recital 166 describes it as a "safety net" for systems which are not considered high risk. In addition, AI systems that are products in areas already covered by harmonised legislation will have to follow these sector-specific regulations.

5.4 Testing without hindering innovation

Alongside obligations, the AI Act also includes measures *"in support of innovation"* (Chapter VI). These measures include AI regulatory sandboxes. The AI Act defines a sandbox as *"a controlled environment that fosters innovation and facilitates the development, training, testing and validation of innovative AI systems for a limited time before their being placed on the market or put into service"* (art. 57.5). Each member state must establish at least one regulatory AI sandbox. Sandboxes serve both to empower businesses to innovate and to foster compliance by allowing stakeholders to learn about regulatory obligations (art. 57.9). This will be particularly useful for SMEs, which have priority access to sandboxes (art. 62.1(a)). In a sandbox environment, providers remain liable for damages, but will not be prosecuted for breaching the AI Act (art. 57.12).

The EU Commission will provide technical support and advice on the establishment and operation of sandboxes. In particular, the AI sandboxes will submit an annual report to the Commission[52]. Further details on the operation of AI sandboxes will be provided in Commission implementing acts (art. 58.1).

But real-world testing can also take place outside AI sandboxes, particularly when testing high-risk AI systems (art. 60.1). These testings can only take place after a real-world testing plan has been approved by market surveillance authorities[53] (art. 60.4) and after the consent of the subjects of testing has been obtained prior to their participation (art. 61.1).

5.5 A new governance ecosystem

To implement and enforce requirements, the AI Act rely on the market surveillance scheme within the meaning of Regulation 2019/1020 (European Parliament and Council, 2019), but also create a brand new ecosystem. As such, the AI Act provides for new bodies, to ensure that the law is properly implemented. Together with existing entities, they should create a European AI governance ecosystem and make sure that legal requirements are met. This choice is justified by recital 148 which stipulates that the governance framework should allow to *"coordinate and support the application of this Regulation at national level, as well as build capabilities at Union level and integrate stakeholders in the field of AI"*. This ecosystem is represented in Figure 6.

[52]Specifically to the AI Office and the AI Board. See below for more information on these entities.
[53]See next paragraph for more information on these entities.

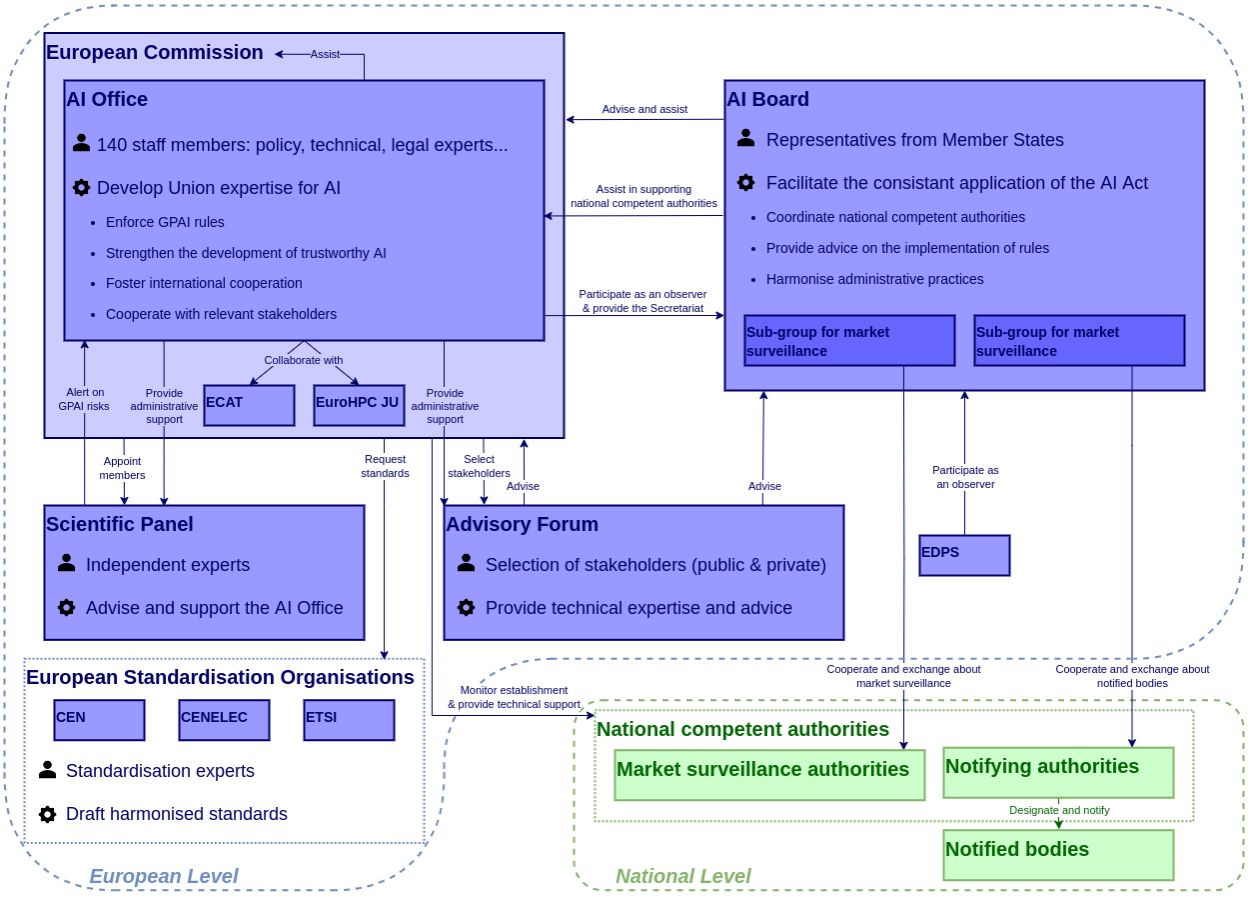

Figure 6: Representation of the European AI governance ecosystem.

National governance

On national level, national competent authorities, consisting of at least one market surveillance authority and one notifying authority, must be designated by each Member State (art. 70.1 AI Act).

Market surveillance authorities are in charge of supervising the placing on the market of AI systems. They act as a point of contact and interface between the public and the Union. For systems covered by existing NLF legislation, existing authorities are automatically designated to also monitor AI systems in their sector. They lead all investigation operations in their sector, with the only exception of GPAI models, which are mainly monitored by EU governance bodies. For these investigations, market surveillance authorities may be granted access to the source code under certain conditions (art. 74.12). Together with the Commission, these market surveillance authorities can also propose *"joint investigations"* to promote compliance or identify non-compliance (art. 74.11). Finally, they can require suppliers to take corrective action in the event of regulatory non-compliance and take action if suppliers refuse to comply (art. 79.5). The AI Act also provides a mechanism for any natural or legal person to lodge a complaint with their market surveillance authority if they believe there has been an infringement of the AI Act (art. 85).

On the other hand, notifying authorities designate and notify conformity assessment bodies to become *"notified body"*, as well as monitor their activities. Notified bodies are responsible for carrying out mandatory conformity assessments for AI systems that require them. They are at the centre of the compliance assessment mechanism for high-risk AI systems put in place by the AI Act.

AI Office

At European level, the first of the new entities created by the AI Act is the AI Office, set up within the European Commission[54] to *"develop Union expertise on AI"* (art. 64.1 AI Act). The AI Office was established by the European Commission Decision of 24 January 2024 (European Commission, 2024b), hereafter EC Dec. The AI Office is notably tasks to enforce the rules on GPAI (art. 2.1 EC Dec) but it also has other missions, such as encouraging innovation or fostering cooperation at different levels (art. 2.2 EC Dec).

To ensure compliance with GPAI rules, the AI Office will be specifically tasked with developing tools and benchmarks for evaluating GPAI models. It will also have a monitoring role, particularly with regard to the emergence of new risks, and the correct implementation of GPAI requirements. For instance, it will conduct evaluation of GPAI models and investigate potential infringements on GPAI rules (art. 3.1 EC Dec). But the AI Office will also be responsible for drafting codes of practice for GPAI models, technical specifications that will help stakeholders comply with the AI Act's GPAI requirements, pending the development of harmonised standards (art. 56.1 AI Act). Providers of GPAI models will be invited to participate in the drawing-up of codes of practice (art. 56.3 AI Act).

The AI Office will also play an important role in assisting the EU Commission, by preparing Commission Decisions, implementing acts and delegated acts. It will also oversee the proper development of standards, prepare standardisation requests and common specifications if necessary. Finally, it will prepare guidance and guidelines in support of the AI Act and provide advice on the implementation of AI sandboxes and real world testing with national competent authorities (art. 3.2 EC Dec). The AI Office will also keep a list of planned and existing AI sandboxes (art. 57.15 AI Act).

[54]More specifically, within the Directorate-General for Communication Networks, Content and Technology (DG CONNECT).

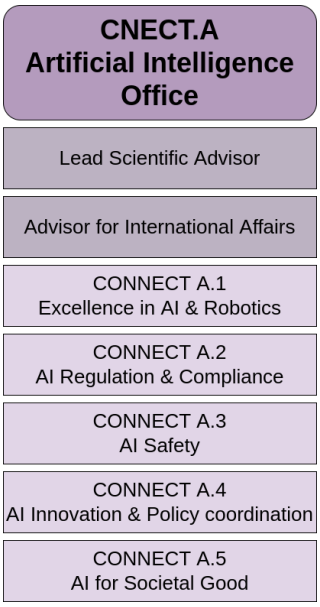

Figure 7: Structure of the AI Office, adapted from (European Commission, n.d.c). All units are independent, there is no hierarchy between units.

Finally, the AI Office will cooperate with other entities: firstly, within the Commission by working with other services (art. 5 EC Dec)[55]; secondly, at international level by supporting other similar institutions or agreements (art. 7 EC Dec); and thirdly, with expert stakeholders, including the industry (art. 4 EC Dec). This last point includes overseeing the AI Pact, an initiative to promote the industry's voluntary commitment to the AI Act requirements ahead of the legal deadline in order to anticipate its impact (European Commission, n.d.b).

As regard to its inner structure, the AI Office will be separated into five units: "Excellence in AI and robotics", "AI Regulation and Compliance", "AI Safety", "AI Innovation and Policy Coordination", and "AI for Societal Good", called CONNECT.A.1 to CONNECT.A.5. They will be supported by a "Lead Scientific Advisor" and an "Advisor for International Affairs". This structure is presented in Figure 7. The structure of the AI Office is in fact a reorganisation of Unit A of the European Commission's Directorate General of Communications Networks, Content and Technology (DG CONNECT). The changes between DG CONNECT A and the AI Office will not require a *"huge reorganisation"*, but the task force will be improved as the AI Office plans to recruit more than 80 people over the next two years, bringing the total workforce over 140 (Gkritsi, 2024).

AI Board

But the AI Office will not be alone in monitoring the correct implementation of the AI Act. The text also provides for the creation of an AI Board, hereafter "the Board", composed of one representative per Member States. The European Data Protection Supervisor will also participate as an observer (art. 65.2 AI Act). The AI Office will attend the Board's meetings without taking parts in the votes (art. 65.2) and provide Secretariat for the Board (art. 65.8). The Board is

[55]Notably the European Centre for Algorithmic Transparency (ECAT) in charge of developing Union expertise for large online platforms and enforce the DSA (art. 5.2(a) EC Dec); or the European High Performance Computing Joint Undertaking (EuroHPC JU), an initiative with private actors to develop a supercomputing ecosystem in Europe (art. 2.3(c) EC Dec).

tasked with supervising that the AI Act is applied consistently in all Member States. This includes coordinating national authorities, providing advice on the implementation of rules and monitor the harmonising of practices (art. 66(a),(c),(d)). The Board will also issue recommendations at the request of the Commission, in particular on existing standards and their use, as well as on common specifications (art. 66(e)(iii),(iv)).

In practice, a large part of the Board's work will take place in thematic subgroups. The AI Act provides for the creation of a subgroup on market surveillance and another on notified bodies, but members of the Board may suggest the creation of new subgroups (art. 65.6). Recently, journalists reported that other subgroups were already in the making, in particular on technical standards, GPAI, innovation and regulatory sandboxes, prohibited systems, high-risk categorisation, or on the interplay with other EU legislation (Bertuzzi, 2024).

Advisory Forum and Scientific Panel

The AI Act also establishes two auxiliary entities: (i) the Advisory Forum, composed of a selection of stakeholders; and (ii) the Scientific Panel, composed of independent experts. The stakeholders of the Advisory Forum and the experts of the Scientific Panel are both appointed by the Commission, but while the experts of the Scientific Panel should be independent from any AI system provider, the Advisory Forum is composed of a wide range of stakeholders, including industry, Small and Medium-sized Enterprises (SMEs), startups, academia and civil society (art. 67.2). Relevant EU agencies and European Standardisation Organisations are also members of the Advisory forum (art. 67.5).

The purpose of the Advisory Forum is to provide technical expertise and advice to the Board and the Commission. For instance, the Commission will consult the Advisory Forum before drafting standardisation requests (art. 40.2) and common specifications (art. 41.1(b)§2). The Scientific Panel is tasked with working with the AI Office, raising alert on GPAI models that could be considered with systemic risk (art. 90), helping with the drafting of codes of conducts and supporting the monitoring activities of the AI Office. Experts from the scientific groups may be called upon by the Commission to carry out the evaluation of GPAI models (art. 92.2). They are also available to Member States requiring expert advice (art. 69).

The objectives of the two bodies are therefore strictly different: whereas the Advisory Forum is designed as a platform for stakeholders to express their interests and raise concerns about the implementation of sector-specific measures, the Scientific Panel is supposed to be independent, impartial and objective (art. 68.4).

5.6 Various deliverables are meant to support the legal requirements

The EU governance ecosystem put in place for AI has an important role to play in the proper implementation of the AI Act and various entities will be involved in developing deliverables which will help support the legal text.

Harmonised standards, developed by the European standardisation bodies at the request of the European Commission, will be the most important of these deliverables. They are expected to define the technical requirements for specifying the legal obligations of the AI Act. They will have direct legal effects and, as such, will be the preferred means of compliance with the requirements set out by the AI Act for high-risk AI systems and GPAI models. As of now, they are mainly drafted for high-risk AI systems, but the Commission may also request standards for GPAI models in the near future.

For the other obligations set out in the AI Act, the European institutions have made provisions for other means of compliance, albeit less powerful than harmonised standards. In particular, codes

of practice are supposed to provide elements of compliance for GPAI models providers (art. 53.4). They do not benefit from the advantages of harmonised standards but may nevertheless be approved by the European Commission by means of implementing acts, in order to give them general validity in the EU (rec. 117). However, they are only intended to supplement the absence of harmonised standards, which would take precedence once published (art. 55.2). Codes of practice will be drawn up by a group of relevant stakeholders, with the assistance of the AI Office and the support of the scientific panel. The AI Office and the AI Board will monitor their implementation (art. 56.1). The stakeholders responsible for drafting these codes of practice should include, in particular, providers of GPAI models and national competent authorities (art. 56.3). Four working groups are currently responsible for drafting the GPAI codes of practice[56], with chairs from different areas of expertise (European Commission, 2024e), selected by the AI Office. After a multi-stakeholders consultation on the codes of practice in the summer of 2024 (European Commission, 2024a), the first meeting of the working groups took place on 30 September 2024, with about a thousand participants (European Commission, 2024d). Discussions are expected to last until 30 April 2025, when a final draft should be presented[57] (European Commission, 2024e).

Codes of conduct are another deliverable that will complement the AI Act. Codes of conduct are voluntary frameworks adopted by providers of non-high-risk AI systems and encouraged by Member States to advance AI literacy (rec. 20). Their development is less stringent and they could emanate from different stakeholders, including providers or deployers of AI systems, although the Commission may also contribute to the development of such initiatives, in particular through the AI Office (art. 95.3). However, the AI Act specifies that such codes should nevertheless be *"developed in an inclusive way"*, with the help of relevant stakeholders, including from civil society and academia (rec. 165).

In addition to codes of practice and codes of conduct, the Commission is empowered to adopt delegated and implementing acts, and to adopt guidelines, with the supervision of the AI Board. Delegated acts will mainly modify requirements of the AI Act, for example by amending the annexes or the conditions for a system to fall into a specific category. Guidelines, on the other hand, will clarify certain requirements of the AI Act, such as when a system should be considered in a certain risk category or how to apply certain provisions of the text. Finally, implementing acts will allow for the approval of existing frameworks or frameworks developed by the Commission itself. In particular, they will be used to approve a code of practice for the transparency obligations of certain AI systems (art. 50.7) and for the obligation of GPAI models (art. 56.6). They will also be used to establish "common specifications" in the absence of adequate harmonised standards (art. 41.1) and "common rules" in the absence of adequate codes of conduct (art. 50.7 and 56.9). We provide an organised list of what delegated acts, guidelines and implementing acts may contain in Annex B.

5.7 Entry into force

After its publication in the OJEU, the text will now be implemented through several steps, represented in Figure 8.

Indeed, the AI Act first came into force on 1 August 2024 but many requirements will apply later. Prohibitions on unacceptable risk systems will apply from 2 February 2025, obligations for GPAI models will apply from 2 August 2025 and transparency obligations from certain AI systems will apply from 2 August 2026. For high-risk systems, requirements will apply from 2 August 2026

[56]These are "Transparency and copyrighted-related rules", "Risk identification and assessment, including evaluations", "Technical risk mitigation", "Internal risk management and governance of general-purpose AI providers"

[57]The strict deadline imposed by the AI Act is 2 May 2025 (art. 53.9).

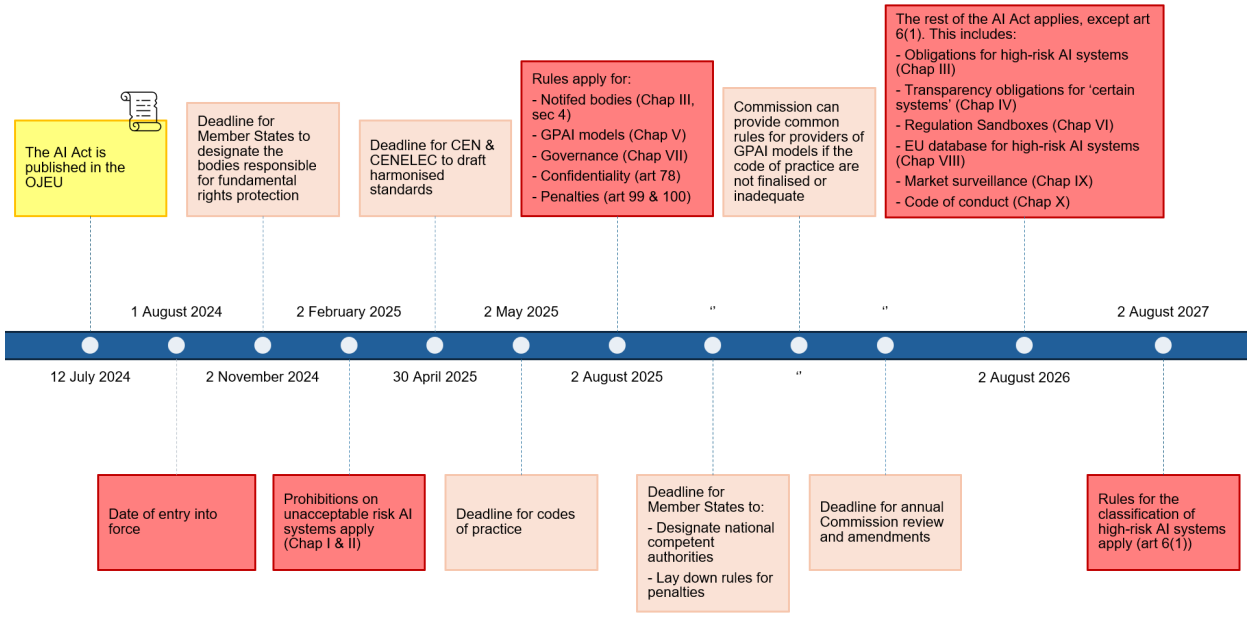

Figure 8: Timeline of implementation of the AI Act.

for systems listed in Annex III but only from 2 August 2027 for systems covered by harmonised legislation listed in Annex I.

6 Criticising the AI Act: what scope, obligations and enforcement mechanisms?

When the Commission first proposed the draft AI Act in 2021, the text attracted a large number of criticisms. Some of them have been addressed in the final version, but the most structural criticisms remain unchanged. In this Section, we briefly go through these criticisms[58].

6.1 A complicated agreement on the definition of AI systems

The definition of an AI system has evolved significantly since the European Commission's first proposal in April 2021. Notably, in the Commission's proposal (European Commission, 2021e), the definition of AI systems was left essentially to Annex I, which, at the time, contained a list of three types of approaches which could be considered AI: (a) Machine learning approaches, (b) Logic- and knowledge-based approaches, and (c) Statistical approaches and optimisation methods.

The first two approaches refer to the two main families of AI, while the third encompasses certain computer programs not normally considered AI. Scholars have pointed out that it was too broad a definition, likely to give rise to legal uncertainty (Ruschemeier, 2023). Some even argued that only machine learning systems should be regulated by the AI Act. In their views, this broad scope was justified for AI systems that present unacceptable risks, as the ban of these systems is justified by the dangers they pose to society and individuals regardless of the technology utilised. However, these critics believed that obligations for high-risk systems were based on characteristics specific to

[58]It should be noted that this list of criticisms is not exhaustive. Interested readers can consult the references cited, which generally provide other points of criticism in addition to those mentioned here.

ML systems, such as opacity or dependency on data (Ebers et al., 2021) and that encompassing other systems in the scope of these obligations would lead to overregulation.

In the final version of the text, the definition adopted does not refer directly to technical terms, and emphasises instead the autonomy of these systems. This definition is much closer to the definition given by the Organisation for Economic Co-operation and Development (OECD)[59]. According to some scholars, this definition further broadens the scope of the AI Act, moving towards a definition of software rather than AI (Hacker, 2024). The clarifications provided by the various recitals draw a clearer line between simple software and AI, but they still leave gaps, as it is not clear whether statistical approaches are taken into account. It will be up to case law to define the frontiers of what is considered AI and what is not.

6.2 What systems should be regulated?

Criticism over the AI Act focuses primarily on the scope of the legal text, i.e. the exact nature of the systems regulated. According to Smuha et al. (2021) the list of prohibited systems is too restrictive. For instance, military applications are excluded from the AI Act. For some scholars, this is a significant gap (Smuha et al., 2021), as AI applications for defence purposes raise many ethical and deontological questions, especially as many scholars are now calling for a ban on autonomous weapons systems (Brand, 2022). This omission could also pose problems for systems that have a dual use and can be used for both civil and military purposes (Ruschemeier, 2023). In addition to military and defence purposes, non-professional purposes and systems used solely for research purposes[60] are also excluded from the scope of the AI Act, leaving the door open to potentially harmful systems (Smuha and Yeung, 2024).

Even in systems that are actually included in the list of prohibitions, certain questions remain. The inclusion of subliminal techniques, for example, has left some researchers wondering what it could possibly contain (Ebers et al., 2021). Furthermore, the ban on social scoring in the AI Act is limited to public entities, ignoring the use of such technologies in the private sector, with dangerous applications that are currently outside of the scope of the AI Act, such as credit scoring (Ebers et al., 2021).

One of the most debated application was probably biometrics. Indeed, the story of the inclusion of biometrics in one category or the other has been one of back and forth. Since the amendments from the three European institutions contained various exceptions and inclusions, biometrics and facial recognition in particular was a sensitive topic during the trilogue negotiations (Bertuzzi, 2022), alongside military applications. In particular, some countries, such as France, strongly opposed the inclusion of military applications and pushed for more exceptions on facial recognition. The adopted version finally comes close to the first proposal of the Commission, with a ban on biometrics systems used by law enforcement authorities in specific circumstances[61]. This ban is further accompanied by broad exceptions which, according to some scholars, do not protect individuals against the dangers of these technologies for fundamental rights (Ruschemeier, 2023). Some have proposed to extend the ban to any biometric system used in public spaces (Ebers et al., 2021), or have called for an additional ban on any emotion recognition system (Wachter, 2024), without success.

[59] In its "Recommendation on AI" first adopted in 2019, the OECD defines an AI system as *"a machine-based system that can, for a given set of human-defined objectives, make predictions, recommendations, or decisions influencing real or virtual environments. AI systems are designed to operate with varying levels of autonomy"* (OECD, 2019). This definition was slightly modified in 2024 to get even closer to the definition of the AI Act (OECD, 2024).

[60] On the contrary, some consider that the first version of the AI Act lacked exceptions for research purposes (Ruschemeier, 2023; Ebers et al., 2021). This problem has therefore been resolved in the current final version, which explicitly excludes them.

[61] In public spaces, in real-time.

The list of high-risk AI systems was also criticised, with some scholars considering that it was too restrictive (Smuha et al., 2021) and suggesting applications that could be added, such as AI systems for housing purposes (Ebers et al., 2021). For Edwards (2022), although the Commission retains the right to modify this list in theory, in practice it will probably be difficult to add new systems to the list. Furthermore, the negotiations during the trilogue resulted in the addition of some exceptions to the classification of high-risk AI systems, such as when a system is intended to perform a narrow task or simply complement human activity without replacing it (art. 6.3). This last-minute change to the text has been heavily criticised, as it complicates the assessment of a system's risk category and creates dangerous loopholes (Wachter, 2024).

Finally, in previous versions of the text, the question of open-source AI system was not addressed, leading scholars to wonder whether the obligations will be the same as for other systems (Ebers et al., 2021). The final version of the AI Act clarifies that the text does not apply to open-source systems unless they fall into the category of unacceptable risk, high risk or GPAI (art. 2.12). This broad exception leaves many open-source system applications still regulated. However, providers of open-source high-risk AI systems and of open-source GPAI models which do not present a systemic risk are exempt of some information and documentation obligations (art. 25.4 and art. 53.2).

6.3 What should be required of AI systems?

The requirements for high-risk AI systems were also at the center of debates, with some scholars considering that the provisions not sufficient to protect against the harms generated by AI and ensure protection of fundamental rights (Smuha et al., 2021). For instance, the data governance requirements fail to explain which biases should be mitigated and what types of discrimination are considered (Ebers et al., 2021), as well as how these biases could be mitigated, while the academic literature flourishes on these issues (Wachter, 2024). Wachter (2024) also highlights the absurdity of requiring a dataset to be "representative" in a world where any set of historical data is biased, explaining that *"neutral data is a fantasy"*. As Article 10 is included in one of the Standardisation Request items, these points should be subsequently covered by harmonised standards.

Other requirements of high-risk AI systems were also criticised. For instance, some consider the requirement on human oversight to be impractical, as it is not yet possible to fully understand a system as the article would require, and the AI Act does not specify when oversight is necessary (Ebers et al., 2021).

Furthermore, the transparency requirement does not say any thing on the interpretability of the systems' output (Ebers et al., 2021), although it should be noted that a "right to explanation of individual decision-making" has been added in the final version (art. 86). It provides for the right to obtain from the deployer *"clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken"*. However, this right is only applicable to the high-risk systems listed under Annex III point 1 and point 3 to 8. It is not applicable for AI systems used in critical infrastructures (Annex III point 2), nor high-risk AI systems that already fall under product safety regulations listed in Annex II.

In addition, some worry about the newly added obligation to conduct a FRIA (art. 27). FRIAs were added thanks to the (European Parliament, 2023a, amendment 413), after their absence in the Commission's version of the AI Act was heavily criticised (Edwards, 2022). However, for Hacker (2024), FRIAs are unlikely to be effective, as they are simply a means to tick boxes, rather than genuinely assess fundamental rights. Some also point out that FRIA are only an obligation of deployers and not of providers of high-risk AI systems, that they only apply to deployers that are public entities, acting on behalf of public entities or providing public services (Wachter, 2024). Furthermore, as with the right to an explanation, the provision to conduct a FRIA excludes critical

infrastructures and AI systems that already fall under product safety regulations.

Finally, the absence of certain requirements is also criticised, as the environmental cost of these systems is hardly taken into account (Wachter, 2024; Hacker, 2023).

In addition to facial recognition and military applications, discussions during the trilogue were also heated regarding Generative AI (GenAI) (Bertuzzi, 2023). According to some scholars and journalists, the provisions relating to GPAI models are the result of strong lobbying efforts by GenAI companies such as the French company Mistral AI or the German company Aleph Alpha (Wachter, 2024; Chan, 2023). As a result, the obligations of GPAI models focus more on transparency than on liability, and compliance is assessed through codes of practices rather than hard regulation. These rules on GPAI are seen as *"extremely weak"* by some scholars (Hacker, 2023), who are therefore calling for additional obligations to be added, such as guaranteeing a high level of cybersecurity for GenAI models (Hacker, 2024).

In addition, the 10^{25} FLOPS threshold is likely to cover very few current systems, since the freely available version of ChatGPT and the Mistral and Aleph models are, for instance, below this threshold (Wachter, 2024). Scholars also criticise the choice of a strict threshold which does not necessarily represent a level of danger and remains highly arbitrary (Smuha and Yeung, 2024). Such threshold may encourage providers of GPAI models to remain below the threshold without reducing the harmful effects of their models (Wachter, 2024).

6.4 The use of CE marking

Some scholars have argued that making the AI Act a product safety regulation is a way of playing on the EU's strength (Almada and Petit, 2023). Indeed, these regulations are generally regarded by the EU as great successes (Mazzini and Scalzo, 2023). However, this view is not shared by all experts. The use of technical standards and the CE mark, even outside AI, has already attracted its fair share of criticism in the literature, as there is no guarantee that a product bearing the CE mark will actually be "safe".

This was clearly demonstrated during the PIP scandal, named after the Poly Implant Prothèse company in France. The case came to light in 2010 after the PIP company, known for providing silicone breast implants, was accused of failing to comply with quality standards in the manufacture of some of its implants. At the time, only one type of silicone gel was allowed for breast implants and PIP, in order to cut costs, manufactured some of their implants with a mixture of this gel and a sub-standard industrial silicone gel (van Leeuwen, 2014). The French public control agency received signals that some of the PIP implants were causing health issues to the women bearing them, with an alleged risk of breast cancer, leading to their withdrawal from the market and the liquidation of the company. The implants were certified and CE marked by TÜV Rheinland. Indeed, CE marking for class III medical devices, which include breast implants, requires a third party certification by a notified body (Rott, 2019). The corresponding EU Directive also imposed surveillance duties on the notified body, such as periodically carrying out appropriate inspections to the manufacturer's quality system, and TÜV Rheinland is alleged to have failed[62].

The Pendra case is another example of damage caused by a CE marked product. It concerns a glucose monitoring device that was prematurely marketed by the Pendra company in the Netherlands in 2003. The device was not suitable for a large number of people due to the different properties of the skin and underlying tissues (Wentholt et al., 2005). Although details of the evaluation process

[62]First, national courts of Germany and France reached different conclusions about the scope of the obligations of certification bodies and their possible liability in case of damage and harm (van Leeuwen, 2014). The extent of the notified body's duties was finally addressed in the CJEU case of Elisabeth Schmitt, where the CJEU confirmed notably that the notified body is for instance *"not under a general obligation to carry out unannounced inspections"*.

have not been disclosed[63], some believe that the notified body should have identified this alleged defect.

While products subjected to external controls can be still be defective, this is even more true for products where no third-party audit is required before affixing the CE mark. As such, toy recalls in the EU, which uses CE marking and self-assessment, are ten to twenty times higher than in the US, where toy certification requires independent third-party certification (Larson and Jordan, 2019). As most high-risk AI systems will fall under this self-assessment procedure, some fear that this will not be enough to guarantee the safety of AI systems (Wachter, 2024).

6.5 The difficulties in effectively implementing obligations

Finally, some of the criticisms also relate to the implementation of the regulation in practice. Some are particularly concerned about the lack of institutional strength for effective enforcement[64], or the lack of democratic supervision. Given that an AI system will pass through many hands between the time it is produced and the time it is actually deployed, some also worry about the allocation of responsibilities along the value chain (Edwards, 2022). In addition, the instability of the technology could create a *"pacing problem"* (Marchant, 2011) whereby systems developed after regulation's entry into force will not be properly covered. This was witnessed with the provisions on GPAI, which were added at the last minute, after GenAI boom the year before (Almada and Petit, 2023). Yet, there is no guarantee that such situation will not happen again in the future.

Furthermore, critics worry about the lack of complaint mechanism and the fact that the AI Act does not facilitate recourse (Ebers et al., 2021). However, this issue is partially addressed in the final version of the AI Act, with Article 85 allowing individuals or groups of individuals to lodge a complaint with a market surveillance authority and Article 86 providing for a right to explanation of individual decision-making. Despite the limited scope of these two rights, scholars consider this to be a positive development in the AI Act, moving towards a right to explanation that was virtually non-existent beforehand (Wachter, 2024).

Criticisms are generally addressed to the regulation itself as well as the means of compliance, but there are also concerns over the expected impact for industry and civil society (Vainionpää et al., 2023). Indeed, although a vast majority of the literature calls for more stringent requirements, another part worries about the potential compliance cost. This cost could range from a few thousand euros for the compliance of one AI system, to several hundred thousand euros for setting up a quality management system, as requested by Article 17 (Haataja and Bryson, 2021). However, this cost will only apply to large companies, as micro-entreprises may comply with certain elements of the quality management system in *"a simplify manner"* (art. 17). Nevertheless, SMEs could still suffer from this compliance cost (Hacker, 2024).

7 A distinctive approach to measuring risks and ensuring rights

Despite other criticisms, the aspect of the AI Act that has really been at the center of debate is that of framing the text not according to the rights of individuals, but on a compliance framework that examine a level of risk to fundamental rights. In this Section, we analyse what risk management looks like in the AI Act and why conflating risks and rights could weaken the protection of fundamental rights.

[63] This was seen by some as a lack of transparency that directly affects the consumer's right to information (Wentholt et al., 2005).

[64] This was reinforced in the final version of the text, with the creation of the AI Office, the AI Board and the whole governance ecosystem. For more information, see Section 5.5.

7.1 Risks in the AI Act must be reduced to an "acceptable" level

Following policy documents recommendations and a series of political choices, the AI Act has adopted a risk-based approach to the categorisation of systems. This risk-based structure implies that some applications of AI systems are "acceptable", while others – the "unacceptable risk systems" – are not (Laux et al., 2023).

But this risk-based approach, and its sister notion of acceptability, is not limited to the systems category; it can also be found directly in the requirements of the legal text, particularly in Article 9 on risk management. The risk management requirement is central in the AI Act (Schuett, 2023b). It is the first requirement set out for high-risk systems, and arguably encompasses all the other requirements. According to Article 9, providers of high-risk AI systems must establish, implement, document and maintain a risk management system, comprising notably of the identification of known and foreseeable risks and the adoption of appropriate measures to eliminate or mitigate those risks. Following mitigation measures, residual risks must be reduced to an "acceptable" level.

? examines what "acceptable" risk might mean in the context of the AI Act. In their view, the European Commission encourages risk reduction "as far as possible", i.e. insofar as it is feasible, whatever the costs. The European Parliament's amendments (European Parliament, 2023a) tended instead to encourage risk reduction "as far as reasonably possible", weighing up costs and benefits before deciding on a threshold of acceptability. The final version of the AI Act (European Parliament and Council, 2024) balances these two versions as risks must be eliminated *"as far as technically feasible through adequate design and development of the high-risk AI system"* (art. 9.5(a)). The acceptable level of risk will therefore be dictated mainly by the state of the art.

7.2 The AI Act between risk- and rights-based approaches

This structure around risk and acceptability determined by the state of the art is not new, since the GDPR, for example, followed the same pattern. This time, however, risks are considered in the AI Act as regards to the *"heath, safety and fundamental rights"*. The text also shows the intention of the European institutions to put in place various tools – standards, codes of practice, impact assessments and so on, to identify and prevent potential violations of fundamental rights. But in doing so, manufacturers will have to determine how to measure a level of risk to fundamental rights, as well as a level of acceptability of such a risk. The question is therefore whether it is possible to reconcile a risk-based approach with a more traditional right-based approach.

Risk-based regulatory approaches take their roots in the safety of critical infrastructures, but they have recently been widely applied in the context of digital technology regulation, such as the GDPR, DSA or the AI Act (Maxwell, 2022). They follow the philosophical movement of "utilitarianism", trying to maximise benefits by balancing economic interests and social well-being through quantitative analysis. This calculation takes the form of risk assessments, which help support the regulation, by providing a means to identify, assess and control risk. Risk is then understood as the combination of the probability of occurrence of an harm and its severity[65] Risk-based regulation is generally seen as a flexible and *"functionally efficient"* tool, designed to accompany a culture of risk management within companies (Black, 2010). Rights-based approaches, on the other hand, are not clearly defined in the academic literature but can be considered to be based on fundamental rights and the philosophical movement of "deontology", and place the individual at the center of moral and legal debate[66]. In rights-based regulation, rights are considered non-negotiable and must

[65]This is the definition usually presented in European law such as the AI Act (European Parliament and Council, 2024, art. 3(2)) or the Product Safety Regulation (European Parliament and Council, 2023c), but it is also a widely accepted definition in risk management (Aven, 2016), where the focus is more on business risk.

[66]See in particular the work of Rawls (1971).

be respected regardless of the level of risk (Hidvegi, 2021). As a result, rights-based approaches generally consider that a violation of fundamental rights is not quantifiable, that all violations are reprehensible and that there is no trade-off with economic benefits whereby a risk to these rights would be acceptable (Maxwell, 2022).

However, a rights-based approach to AI is complicated to put in place as it requires effective enforcement mechanisms (Smuha, 2021). Furthermore, different AI technologies may present different issues and should be regulated differently, which is why many scholars have also emphasised the need to regulate AI through risks (Schuett, 2023a). Others believe that AI regulation should be built on the pillars of liberal democratic societies: fundamental rights, the rule of law and democracy (Smuha et al., 2021). Similarly, Ruschemeier (2023) argues that the regulatory efforts for AI should be focused on enhancing the protection of legal rights, in particular by enacting the precautionary principle. In response, the AI Act attempts to accommodate both, mixing the semantics of the rights-based approach with the mechanisms of the risk-based approach. The place of fundamental rights in the AI Act is therefore still being debated (Almada and Petit, 2023), with some believing that it is no more than a marketing tool, and that economic benefits with the improvement of the European internal market are the predominant objective, relegating fundamental rights to second place (Castets-Renard and Besse, 2022).

7.3 The risks of measuring risks

Risk management is based on two assumptions: first, that it is possible to anticipate every risk; second, that for each risk, it is possible to calculate an associated probability and precise magnitude.

However, neither of these two assumptions is true in all cases. The first one because there is no such thing as zero risk. People assessing the risks might miss situations in which a risk could occur. This is particularly true of risks that are specific to a certain minority group. The fewer people affected, the less likely the risk is to be identified. These frameworks therefore render invisible certain kinds of harms suffered by certain groups of individuals (Kaminski, 2022). Risk measurement is inherently unfair today, as there is often insufficient data on certain groups – such as women or ethnic minorities – to properly assess risks. Drug doses or seatbelt strength are measured on a "general" body type that is not representative of every individual (Perez, 2020), and will only work on the statistically largest or most powerful group. The second assumption is based on a measurement paradigm: the idea that any observable phenomenon can be evaluated in a quantifiable way. But risk measures are often approximate and, by hiding behind "scientific facts", can give a false sense of accuracy (Rothstein et al., 2006).

Risk management also implies that the technology will be adopted despite its harm (Kaminski, 2022). As such, risk assessments can sometimes be used to justify a policy decision that was already made rather than truly giving insight on what should truly be done (Rothstein et al., 2006).

These general conclusions about risk frameworks have even greater implications when the frameworks are applied to fundamental rights. Indeed, both the identification and the measurement of the risk to fundamental rights are highly biased depending on who carries out the assessment. Risk measurement, in particular, can only be carried out using proxies, such as algorithmic unfairness for discrimination, and therefore misses out on a large proportion of possible risks.

7.4 Defining a threshold of acceptability for fundamental rights: an impossible task?

Defining a threshold of acceptability is a difficult issue in risk management. For instance, in the case of the safety of a nuclear power plant, the risk of an accident are quantified, the benefits of

nuclear energy and the costs are weighed up, and the plant is allowed to open if the measured risk is below a certain threshold: the "acceptable level of risk" to safety (Fischhoff, 1983). This threshold cannot be zero, because a "zero risk" approach would lead to the total rejection of nuclear activity, which could otherwise provide benefits to society. Compromise are thus necessary. While there is no direct threshold for the safety level of a nuclear power plant, this safety level itself depends on the threshold granted to the various safety components. To take another example, in the case of a car, standardised seat-belt robustness tests are carried out to assess that the risk of dying in a car accident is less than a certain probability.

However, even if such a safety threshold is commonly used in product safety regulations, its adaptation to fundamental rights, such as a discrimination threshold, is not straightforward. For instance, if fairness allow us to measure some notion of discrimination, then residual biases or unfairness must be understood as the expression of residual risks as per the AI Act (Orwat et al., 2024). Setting an acceptable level of risk of discrimination may then involve defining a threshold for fairness metrics. There are, however, many different measures of fairness that can be used in different situations (Barocas et al., 2021), some of them are sometimes even incompatibles (Chouldechova, 2017), and it is impossible to define an universal choice. Indeed, fairness is highly context-dependent (Wachter et al., 2021) and is therefore hard to standardise (Bringas Colmenarejo et al., 2022).

Additionally to these difficulties, setting a threshold for fundamental rights also poses a question of where to draw the line. It is usually admitted that fundamental rights follow logic of optimisation (Alexy, 2010). This means that they must be extended to a maximum and only the least restrictive solution on fundamental rights should be accepted. However, in the AI Act, there is a shift from this "optimising logic" to a "satisfactory logic" whereby any solution above a certain threshold could be deemed acceptable (Almada and Petit, 2023). These two types of logic – optimisation and satisfaction – translates into what Busch (2011) calls "olympics thresholds", i.e. the best possible, and "filter thresholds", i.e. better than a certain limit. If the satisfactory logic may be better in a context of technology uncertainty where the state of the art is dynamic and the "best" solution changes quickly (Almada and Petit, 2023), the optimising logic may better protect fundamental rights.

However, both of these rationales do not account for the fact that setting a threshold can be harmful in itself. Yet, scholars have shown that *"threshold theory"*, i.e. the science of associating a quantifiable level with the acceptability of a harm, is in fact a strategy of assimilation whereby science is used to justify damage (Liboiron, 2021)[67]. According to this view, which is closer to that of the advocates of right-based regulation, rights and wrongs should not be quantified. Nevertheless, this is not the approach taken by the Commission with the AI Act, which assumes that thresholds need to be set for the risks of AI.

7.5 Reconciling rights and risks: a distinctive approach which challenges the operation of the NLF

The distinctive approach to AI regulation led by the EU with its AI Act, mixes risk-based and rights-based regulation and proposes to quantify the unquantifiable, i.e. to set a level of "acceptable" risk to fundamental rights. This dual approach, which was also noted by a number of scholars (Ho-Dac, 2023; Almada and Petit, 2023; Gornet and Maxwell, 2024; Smuha and Yeung, 2024), is due to the fact that the AI Act is product safety regulation where compliance with the state of the art – for instance technical standards – is supposed to ensure a level of protection of the product consumer,

[67] Liboiron (2021) associates threshold theory to colonialism, as thresholds are used to justify the pollution of indigenous lands.

here, the end-user of an AI system. Yet, while these requirements are generally considered to relate to safety or health, the AI Act also aims to protect fundamental rights. The objective of the AI Act –to protect fundamental rights –and the means it implements to achieve it –risk management and product compliance –are therefore not necessarily aligned (Smuha and Yeung, 2024).

For Almada and Petit (2023), the AI Act's approach to regulate through product safety and technical standards necessarily involves a "logic of evaluation", where risks should be kept below a certain threshold. On the contrary, the traditional approach to fundamental rights protection involves a "logic of proportionality", where risks should be minimised as far as possible. Ebers et al. (2021) further notes that the inclusion of individual rights into the AI Act, especially those ex-post such as recourse mechanism, might challenge the approach of the NLF which is fundamentally ex-ante. Standards could be forced to define thresholds, either by setting them at the best known performance at the time the standard was drafted, which risks becoming obsolete very quickly, or at a certain level of acceptability which will necessarily be very arbitrary. Another solution could be to leave room for interpretation for judges to decide what "acceptable" means in a given situation. However, this would require the company to decide beforehand what "acceptable" means for them, with the risk that the two visions might not align.

8 Conclusion

In this work, we have presented the core of the European approach to AI regulation: the AI Act. This complex piece of legislation takes its roots in ethics charters and policy documents, following their advice to adopt a risk-based regulation with a special focus on "trustworthiness", i.e. respecting ethical and legal frameworks while being technically robust.

The risk-based approach was notably tested in the previous major success of European regulation in the digital sector: the GDPR. However, the AI Act goes further by proposing to adopt a product safety approach, whereby different risk categories of systems will have to comply with different legal obligations that will be assessed before the system is put on the market. These categories are defined in particular according to the sector in which the AI system is deployed, and the AI Act establishes a pyramid of these risks, ranging from higher risk with strict requirements, to lower risk with softer requirements. However, this classification is challenged by a cross-sectoral category: GPAI models, and its systematic risk subdivision, which include most of the current GenAI models. To ensure that obligations are met and that the AI Act is enforced, a whole ecosystem was designed to implement the various parts of the text, to enable the evaluation of systems and the deployment of measures in favour of innovation.

If the AI Act faced numerous criticisms when it first came out in April 2021, it was above all its dual approach, mixing risks and rights, that attracted attention the most. Indeed, the AI Act will require high-risk AI systems providers to put in place a risk management system and to reduce risks to a "technically acceptable level", which will depend on the state of the art, in particular the content of technical standards. This risk-based approach for fundamental rights will, however, inevitably pose problems when it comes to defining technical requirements in standards. It remains to be seen how far standards will go in making normative choices such as setting a threshold of acceptability. But even if they do not go that far, the mismatch between a risk-based approach, where compliance is assessed *ex ante* and a rights-based approach, where violations of rights are assessed *ex post*, could challenge the operation of the NLF.

A Legal ecosystem of the AI Act

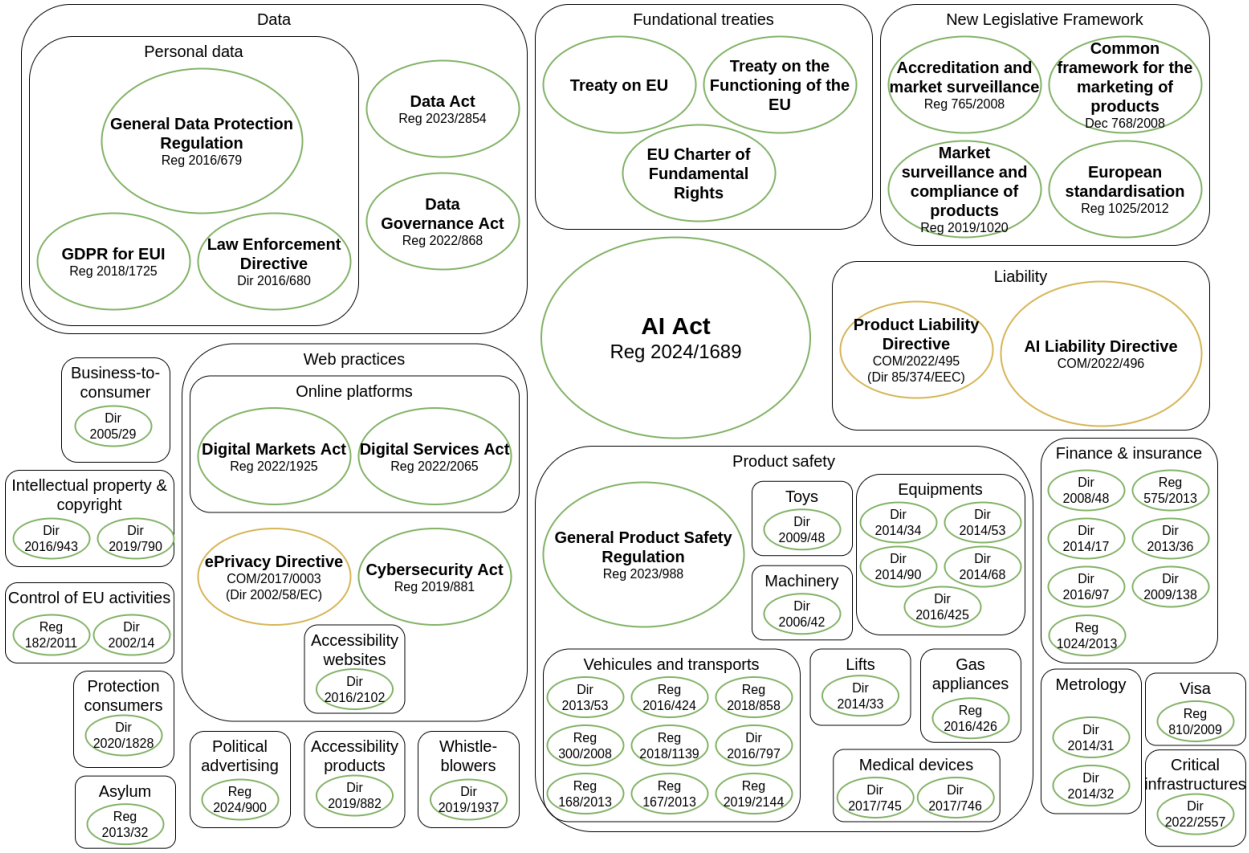

Figure 9: Legal ecosystem of the AI Act. Each Regulation and Directive that appear is cited in the AI Act, yellow bubbles are texts in construction or revision, green bubbles are already voted. The references to the legal texts are organised by the author.

B List of what the European Commission's delegated acts, guidelines and implementing acts may contain in the AI Act

- Delegated acts may be adopted for:
  1. High-risk systems, in order to:
     - amend conditions where an AI system under Annex III is exceptionally not considered high-risk (art. 6.6 and 6.7)
     - amend Annex III to add, modify or remove AI systems from the list (art. 7.1 and 7.3)
     - subject high-risk AI systems under Annex III to a conformity assessment with a notified body (art. 43.6);
     - amend Annex IV on the technical documentation, Annex V on the EU declaration of conformity, Annex VI and VII on conformity assessment (art. 11.3, 47.5 and 43.5);
  2. for GPAI models, in order to

  - modify the threshold above which GPAI models are considered to present a systematic risk (art. 51.3);
  - amend Annex XIII to specify or update criteria by which a GPAI model is considered to present a systematic risk (art. 52.4);
  - amend Annex XI and XII on technical documentation and transparency obligations for GPAI model providers (art. 53.5 and 53.6).

- Guidelines may be used to precise:
  - when AI system in Annex III are not high-risk, and provide a comprehensive list of examples for systems that are considered high-risk and those which are not (art. 6.5);
  - the elements of a quality management system (art. 63.1);
  - how to apply article 3 on the definition of high-risk AI systems, article 5 on prohibited practices, articles 8 to 15 and 25 on the requirements for high-risk AI systems and the responsibilities across the AI value chain, and article 50 on transparency obligations (art. 96.1);
  - the provisions on substantial modification (art. 96.1);
  - the relation between the AI Act and other harmonised legislation (art. 96.1).

- Implementing acts may be used to:
  - approve a code of practice for the transparency obligations of certain AI systems (art. 50.7) and for the obligation of GPAI models (art. 56.6);
  - establish "common specifications" in the absence of adequate harmonised standards (art. 41.1);
  - establish "common rules" in the absence of adequate codes of conduct (art. 50.7 and 56.9).
  - suspend or withdraw the notification of a notified body (art. 37.4);
  - precise arrangements for regulatory sandboxes (art. 58.1) and real-world testing plans (art. 60.1);
  - establish the scientific panel and precise its procedures (art. 68.1 and 68.4);
  - precise the fees for access to the pool of experts (art. 69.2);
  - give a template for post-market monitoring plans (art. 72.3);
  - detail the conditions of evaluations of GPAI models by the AI Office (art. 92.6);
  - precise the procedure for fines (art. 101.6).

References

Ada Lovelace Institute (2021). Three proposals to strengthen the EU Artificial Intelligence Act. https://www.adalovelaceinstitute.org/blog/three-proposals-strengthen-eu-artificial-intelligence-act/

Alexy, R. (2010). A Theory of Constitutional Rights. *Oxford University Press*. https://global.oup.com/academic/product/a-theory-of-constitutional-rights-9780199584239

Almada, M. and Petit, N. (2023). The EU AI act : a medley of product safety and fundamental rights? Technical report, European University Institute. https://hdl.handle.net/1814/75982

appliedAI (2023). AI Act: Risk Classification of AI Systems from a Practical Perspective. Technical report, Initiative for applied artificial intelligence. https://www.appliedai.de/en/insights/ai-act-risk-classification-of-ai-systems-from-a-practical-perspective

Aven, T. (2016). Risk assessment and risk management: Review of recent advances on their foundation. *European Journal of Operational Research*, 253(1):1–13. https://doi.org/10.1016/j.ejor.2015.12.023

Barocas, S., Hardt, M., and Narayanan, A. (2021). *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press. http://www.fairmlbook.org

Bertuzzi, L. (2022). AI Act: EU Parliament's discussions heat up over facial recognition, scope. *Euractiv*. https://www.euractiv.com/section/digital/news/ai-act-eu-parliaments-discussions-heat-up-over-facial-recognition-scope/

Bertuzzi, L. (2023). EU's AI Act negotiations hit the brakes over foundation models. *Euractiv*. https://www.euractiv.com/section/artificial-intelligence/news/eus-ai-act-negotiations-hit-the-brakes-over-foundation-models/

Bertuzzi, L. (2024). EU countries make significant changes to AI Board's procedures, structure. *MLex Market Insight*. https://mlexmarketinsight.com/news/insight/eu-countries-make-significant-changes-to-ai-board-s-procedures-structure

Black, J. (2010). Risk-based Regulation: Choices, Practices and Lessons Being Learnt. In *Risk and Regulatory Policy: Improving the Governance of Risk*, pages 185–224. OECD Reviews of Regulatory Reform. https://doi.org/10.1787/9789264082939-11-en

Brand, J. L. M. (2022). Why reciprocity prohibits autonomous weapons systems in war. *AI and Ethics*, 3(2):619–624. https://doi.org/10.1007/s43681-022-00193-1

Bringas Colmenarejo, A., Nannini, L., Rieger, A., Scott, K. M., Zhao, X., Patro, G. K., Kasneci, G., and Kinder-Kurlanda, K. (2022). Fairness in Agreement With European Values: An Interdisciplinary Perspective on AI Regulation. In *AIES '22: Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, pages 107–118. Association for Computing Machinery. https://doi.org/10.1145/3514094.3534158

Busch, L. (2011). *Standards: Recipes for Reality*. The MIT Press. https://doi.org/10.7551/mitpress/8962.001.0001

Castets-Renard, C. and Besse, P. (2022). Ex ante Accountability of the AI Act: Between Certification and Standardization, in Pursuit of Fundamental Rights in the Country of Compliance. In Castets-Renard, C. and Eynard, J., editors, *Artificial Intelligence Law: Between Sectoral Rules and Comprehensive Regime. Comparative Law Perspectives*. Bruylant. https://papers.ssrn.com/abstract=4203925

Celeste, E. (2019). Digital constitutionalism: a new systematic theorisation. *International Review of Law, Computers & Technology*, 33(1):76–99. https://doi.org/10.1080/13600869.2019.1562604

CEN (n.d.). CEN/CLC/JTC 13 – Cybersecurity and Data Protection. Published standards. EN 17529:2022. https://standards.cencenelec.eu/dyn/www/f?p=CEN:110:0::::FSP_PROJECT,FSP_ORG_ID:63633,2307986&cs=11F702120AA40D5CC2DD42848140B1806

Chan, K. (2023). Europe's world-leading artificial intelligence rules are facing a do-or-die moment. *Quartz*. https://qz.com/europes-world-leading-artificial-intelligence-rules-are-1851069721

Chouldechova, A. (2017). Fair Prediction with Disparate Impact: A Study of Bias in Recidivism Prediction Instruments. *Big Data*, 5(2):153–163. Mary Ann Liebert, Inc. publishers. https://doi.org/10.1089/big.2016.0047

CNIL (2020). ISO 27701, an international standard addressing personal data protection. https://www.cnil.fr/en/iso-27701-international-standard-addressing-personal-data-protection

De Gregorio, G. (2021). The rise of digital constitutionalism in the European Union. *International Journal of Constitutional Law*, 19(1):41–70. https://doi.org/10.1093/icon/moab001

Ebers, M., Hoch, V. R. S., Rosenkranz, F., Ruschemeier, H., and Steinrötter, B. (2021). The European Commission's Proposal for an Artificial Intelligence Act – A Critical Assessment by Members of the Robotics and AI Law Society (RAILS). *J*, 4(4):589–603. https://doi.org/10.3390/j4040043

EDPB (2022a). EDPB adopts "wish list"' of procedural aspects, first EU data protection seal and a statement on digital euro. European Data Protection Board. https://www.edpb.europa.eu/news/news/2022/edpb-adopts-wish-list-procedural-aspects-first-eu-data-protection-seal-and-statement_en

EDPB (2022b). Opinion 28/2022 on the Europrivacy criteria of certification regarding their approval by the Board as European Data Protection Seal pursuant to Article 42.5 (GDPR). European Data Protection Board. https://www.edpb.europa.eu/our-work-tools/our-documents/opinion-board-art-64/opinion-282022-europrivacy-criteria-certification_en

Edwards, L. (2022). Expert opinion. Regulating AI in Europe: four problems and four solutions. Technical report, Ada Lovelace Institute. https://www.adalovelaceinstitute.org/wp-content/uploads/2022/03/Expert-opinion-Lilian-Edwards-Regulating-AI-in-Europe.pdf

European Commission (2012a). Charter of Fundamental Rights of the European Union. OJ C 326, p.391–407. http://data.europa.eu/eli/treaty/char_2012/oj/eng

European Commission (2012b). Consolidated version of the Treaty on European Union. OJ C326, p.13–390. http://data.europa.eu/eli/treaty/teu_2012/oj

European Commission (2012c). Consolidated version of the Treaty on the Functioning of the European Union. OJ C 326, p.47–390. http://data.europa.eu/eli/treaty/tfeu_2012/oj/eng

European Commission (2016). Commission Staff Working Document on the implemetation of the actions foreseen in the 2015 and 2016 Union work programmes for European standardisation, including the implementing acts and mandates sent to the European standardisation organisations. Accompanying the document Communication from the Commission to the European Parliament, the Council and the European Economic and Social Committee. The Annual Union work programme for European standardisation for 2016. SWD/2015/0301 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52015SC0301

European Commission (2018a). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Artificial Intelligence For Europe. COM/2018/237 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:237:FIN

European Commission (2018b). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions – Coordinated Plan on Artificial Intelligence. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52018DC0795

European Commission (2019). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Building Trust in Human Centric Artificial Intelligence. COM(2019)168. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52019DC0168

European Commission (2020a). Report from the Commission to the European Parliament, the Council and the European Economic and Social Committee – Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics. COM/2020/64 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0064

European Commission (2020b). White Paper on Artificial Intelligence: a European approach to excellence and trust - European Commission. https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en

European Commission (2021a). Artificial intelligence – ethical and legal requirements. https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-intelligence-ethical-and-legal-requirements/feedback_en?p_id=24212003

European Commission (2021b). Commission Staff Working Document – Impact Assessment Accompanying the Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. SWD/2021/84 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021SC0084

European Commission (2021c). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions – Fostering a European approach to Artificial Intelligence. https://eur-lex.europa.eu/legal-content/en/TXT/?uri=COM:2021:205:FIN

European Commission (2021d). Coordinated Plan on Artificial Intelligence 2021 Review. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/library/coordinated-plan-artificial-intelligence-2021-review

European Commission (2021e). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and Amending certain Union legislative acts. COM/2021/206 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

European Commission (2024a). AI Act: Have Your Say on Trustworthy General-Purpose AI. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/consultations/ai-act-have-your-say-trustworthy-general-purpose-ai

European Commission (2024b). Commission Decision of 24 January 2024 establishing the European Artificial Intelligence Office. OJ C, C/2024/1459. http://data.europa.eu/eli/C/2024/1459/oj

European Commission (2024c). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on boosting startups and innovation in trustworthy artificial intelligence. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2024:28:FIN

European Commission (2024d). The kick-off Plenary for the General-Purpose AI Code of Practice took place online. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/news/kick-plenary-general-purpose-ai-code-practice-took-place-online

European Commission (2024e). Meet the Chairs leading the development of the first General-Purpose AI Code of Practice. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/news/meet-chairs-leading-development-first-general-purpose-ai-code-practice

European Commission (n.d.a). AI Act. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

European Commission (n.d.b). AI Pact. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/ai-pact

European Commission (n.d.c). European AI Office. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/policies/ai-office

European Commission (n.d.d). New legislative framework. *Internal Market, Industry, Entrepreneurship and SMEs*. https://single-market-economy.ec.europa.eu/single-market/goods/new-legislative-framework_en

European Commission (n.d.e). Sectorial AI Testing and Experimentation Facilities under the Digital Europe Programme. Shaping Europe's digital future. https://digital-strategy.ec.europa.eu/en/activities/testing-and-experimentation-facilities

European Council (1985). Consolidated text: Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products. http://data.europa.eu/eli/dir/1985/374/1999-06-04

European Council (2020). Special meeting of the European Council (1 and 2 October 2020) – Conclusions. https://www.consilium.europa.eu/en/press/press-releases/2020/10/02/european-council-conclusions-1-2-october-2020/

European Council (2021). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Presidency compromise text. 14278/21. https://data.consilium.europa.eu/doc/document/ST-14278-2021-INIT/en/pdf

European Council (2022a). Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and Amending certain Union legislative acts - General approach. https://data.consilium.europa.eu/doc/document/ST-14954-2022-INIT/en/pdf

European Council (2022b). Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union - Text de compromis de la présidence - Article 3, paragraphe 1 ter, Articles 4 bis à 4 quater, Annexe VI (3) et (4), considérant 12 bis bis. https://artificialintelligenceact.eu/wp-content/uploads/2022/05/AIA-FRA-Art-34-13-May.pdf

European Council (2022c). Proposition de Règlement du Parlement européen et du Conseil établissant des règles harmonisées concernant l'intelligence artificielle (législation sur l'intelligence artificielle) et modifiant certains actes législatifs de l'Union - Text de compromis de la présidence - Version consolidée. https://data.consilium.europa.eu/doc/document/ST-10069-2022-INIT/x/pdf

European Council (2024). Artificial intelligence (AI) act: Council gives final green light to the first worldwide rules on AI. https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/

European Parliament (2017). European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

European Parliament (2020a). European Parliament resolution of 20 October 2020 with recommendations to the Commission on a civil liability regime for artificial intelligence (2020/2014(INL)). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0276_EN.html

European Parliament (2020b). European Parliament resolution of 20 October 2020 with recommendations to the Commission on a framework of ethical aspects of artificial intelligence, robotics and related technologies (2020/2012(INL)). https://www.europarl.europa.eu/doceo/document/TA-9-2020-0275_EN.html

European Parliament (2023a). Artificial Intelligence Act. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). P9_TA(2023)0236. https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.pdf

European Parliament (2023b). Artificial Intelligence Act: deal on comprehensive rules for trustworthy AI. https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai

European Parliament (2024). Artificial Intelligence Act: MEPs adopt landmark law. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law

European Parliament (n.d.a). Framework of ethical aspects of artificial intelligence, robotics and related technologies. *Legislative Train Schedule*. https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-ai-ethical-framework

European Parliament (n.d.b). List of legal bases providing for the ordinary legislative procedure in the Treaty of Lisbon. https://www.europarl.europa.eu/cmsdata/198025/List_of_legal_bases.pdf

European Parliament and Council (2009). Directive 2009/48/EC of the European Parliament and of the Council of 18 June 2009 on the safety of toys. OJ L 170, p.1-37. http://data.europa.eu/eli/dir/2009/48/oj/eng

European Parliament and Council (2014). Directive 2014/53/EU of the European Parliament and of the Council of 16 April 2014 on the harmonisation of the laws of the Member States relating to the making available on the market of radio equipment and repealing Directive 1999/5/EC. https://eur-lex.europa.eu/eli/dir/2014/53/oj

European Parliament and Council (2016a). Directive (EU) 2016/680 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA. OJ L 119, p.89-131. https://eur-lex.europa.eu/eli/dir/2016/680/2016-05-04

European Parliament and Council (2016b). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). OJ L 18, p.1–88. http://data.europa.eu/eli/reg/2016/679/2016-05-04/eng

European Parliament and Council (2017a). Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on Privacy and Electronic Communications). COM/2017/010 final. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52017PC0010

European Parliament and Council (2017b). Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC. OJ L 117, p.1-175. http://data.europa.eu/eli/reg/2017/745/oj/eng

European Parliament and Council (2018). Regulation (EU) 2018/1725 of the European Parliament and of the Council of 23 October 2018 on the protection of natural persons with regard to

the processing of personal data by the Union institutions, bodies, offices and agencies and on the free movement of such data, and repealing Regulation (EC) No 45/2001 and Decision No 1247/2002/EC. OJ L 295, p.39-98. https://eur-lex.europa.eu/eli/reg/2018/1725/oj

European Parliament and Council (2019). Regulation (EU) 2019/1020 of the European Parliament and of the Council of 20 June 2019 on market surveillance and compliance of products and amending Directive 2004/42/EC and Regulations (EC) No 765/2008 and (EU) No 305/2011. OJ L 169, p.1–44. http://data.europa.eu/eli/reg/2019/1020/oj

European Parliament and Council (2022). Proposal for a Directive of the European Parliament and of the Council on adapting non-contractual civil liability rules to artificial intelligence (AI Liability Directive). https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex:52022PC0496

European Parliament and Council (2022). Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act). http://data.europa.eu/eli/reg/2022/1925/oj/eng

European Parliament and Council (2022a). Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act). OJ L 277, p.1–102. https://eur-lex.europa.eu/eli/reg/2022/2065/oj

European Parliament and Council (2022b). Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (Data Governance Act). OJ L 152, p.1–44. http://data.europa.eu/eli/reg/2022/868/oj/eng

European Parliament and Council (2023a). Regulation (EU) 2023/1230 of the European Parliament and of the Council of 14 June 2023 on machinery and repealing Directive 2006/42/EC of the European Parliament and of the Council and Council Directive 73/361/EEC. OJ L 165, p.1-102. https://eur-lex.europa.eu/eli/reg/2023/1230/oj

European Parliament and Council (2023b). Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828 (Data Act). OJ L. https://eur-lex.europa.eu/eli/reg/2023/2854

European Parliament and Council (2023c). Regulation (EU) 2023/988 of the European Parliament and of the Council of 10 May 2023 on general product safety, amending Regulation (EU) No 1025/2012 of the European Parliament and of the Council and Directive (EU) 2020/1828 of the European Parliament and the Council, and repealing Directive 2001/95/EC of the European Parliament and of the Council and Council Directive 87/357/EEC. OJ L 135, p.1–51. http://data.europa.eu/eli/reg/2023/988/oj

European Parliament and Council (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). OJ L. http://data.europa.eu/eli/reg/2024/1689/oj

Europrivacy (n.d.a). Audit and Certification in Data Protection. https://www.europrivacy.org/

Europrivacy (n.d.b). Europrivacy Benefits and Advantages. https://www.europrivacy.org/en/ep/benefits

Fischhoff, B. (1983). "Acceptable Risk": The Case of Nuclear Power. *Journal of Policy Analysis and Management*, 2(4):559–575. Wiley, Association for Public Policy Analysis and Management. https://doi.org/10.2307/3323574

Future of Life Institute (n.d.). Historic Timeline – EU Artificial Intelligence Act. https://artificialintelligenceact.eu/developments/

Gellert, R. (2018). Understanding the notion of risk in the General Data Protection Regulation. *Computer Law & Security Review*, 34(2):279–288. https://www.doi.org/10.1016/j.clsr.2017.12.003

Gkritsi, E. (2024). Commission's DG for technology restructures to realise AI Office. *www.euractiv.com*. https://www.euractiv.com/section/artificial-intelligence/news/commissions-dg-for-technology-restructures-to-realise-ai-office/

Gornet, M., Delarue, S., Boritchev, M., and Viard, T. (2024). Mapping AI ethics: a meso-scale analysis of its charters and manifestos. In *FAccT '24: Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 127–140, Rio de Janeiro, Brazil. Association for Computing Machinery. https://doi.org/10.1145/3630106.3658545

Gornet, M. and Maxwell, W. (2024). The European approach to regulating AI through technical standards. *Internet Policy Review*, 13(3). https://doi.org/10.14763/2024.3.1784

Grafenstein, M. v. (2022). Co-regulation and competitive advantage in the GDPR: Data protection certification mechanisms, codes of conduct and data protection-by-design. In González Fuster, G., van Brakel, R., and De Hert, P., editors, *Research Handbook on Privacy and Data Protection Law*, Law 2022, pages 402–432. Edward Elgar Publishing. https://doi.org/10.4337/9781786438515

Haataja, M. and Bryson, J. J. (2021). What costs should we expect from the EU's AI Act? https://osf.io/8nzb4

Hacker, P. (2023). What's Missing from the EU AI Act. *Verfassungsblog*. Verfassungsblog. https://www.doi.org/10.59704/3f4921d4a3fbeeee

Hacker, P. (2024). Comments on the Final Trilogue Version of the AI Act. https://dx.doi.org/10.2139/ssrn.4757603

Hidvegi, F. (2021). The EU should regulate AI on the basis of rights, not risks. *Access Now*. https://www.accessnow.org/eu-regulation-ai-risk-based-approach/

HLEG (2019a). Ethics guidelines for trustworthy AI. Technical report, Independent Hight-Level Expert Group on Artificial Intelligence set up by the European Commission. Publications Office of the European Union, Directorate-General for Communications Networks, Content and Technology. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

HLEG (2019b). Policy and investment recommendations for trustworthy Artificial Intelligence. Technical report, Independent Hight-Level Expert Group on Artificial Intelligence set up by the European Commission. Publications Office of the European Union. https://data.europa.eu/doi/10.2759/465913

HLEG (2020a). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self assessment. Technical report, Independent Hight-Level Expert Group on Artificial Intelligence set up by the European Commission. Publications Office of the European Union, Directorate-General for Communications Networks, Content and Technology. https://data.europa.eu/doi/10.2759/002360

HLEG (2020b). Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence. Technical report, Independent Hight-Level Expert Group on Artificial Intelligence set up by the European Commission. Publications Office of the European Union, Directorate-General for Communications Networks, Content and Technology. https://data.europa.eu/doi/10.2759/733662

Ho-Dac, M. (2023). Considering Fundamental Rights in the European Standardisation of Artificial Intelligence: Nonsense or Strategic Alliance? In Jakobs, K., editor, *Joint Proceedings EURAS & SIIT 2023 – (Responsible) Standardisation for Smart Systems*. Verlag Mainz. https://hal.science/hal-04411136

ISO/IEC (2005). ISO/IEC 27001:2005, Information technology — Security techniques — Information security management systems — Requirements. Edition 1. https://www.iso.org/contents/data/standard/04/21/42103.html

ISO/IEC (2019). ISO/IEC 27701:2019, Security techniques — Extension to ISO/IEC 27001 and ISO/IEC 27002 for privacy information management — Requirements and guidelines. Edition 1. https://www.iso.org/standard/71670.html

ISO/IEC (2022a). ISO/IEC 27001:2022, Information security, cybersecurity and privacy protection — Information security management systems — Requirements. Edition 3. https://www.iso.org/standard/27001

ISO/IEC (2022b). ISO/IEC 27002:2022, Information security, cybersecurity and privacy protection — Information security controls. Edition 3. https://www.iso.org/standard/75652.html

ITEH Standards (n.d.). Mandate M/530 - Privacy Management. https://standards.iteh.ai/catalog/mandate/cen/90507928-bb87-4b60-8c56-fdd18d6dd2db/m-530

Kamara, I. (2017). Co-regulation in EU personal data protection: the case of technical standards and the privacy by design standardisation 'mandate'. *European Journal of Law and Technology*, 8(1). https://ejlt.org/index.php/ejlt/article/view/545

Kaminski, M. E. (2022). Regulating the Risks of AI. *Boston University Law Review, U of Colorado Law Legal Studies*, 103(1347). https://dx.doi.org/10.2139/ssrn.4195066

Lachaud, E. (2018). The General Data Protection Regulation and the rise of certification as a regulatory instrument. *Computer Law & Security Review*, 34(2):244–256. https://doi.org/10.1016/j.clsr.2017.09.002

Lachaud, E. (2020). What GDPR tells about certification. *Computer Law & Security Review*, 38. https://doi.org/10.1016/j.clsr.2020.105457

Larson, D. B. and Jordan, S. R. (2019). Playing it safe: toy safety and conformity assessment in Europe and the United States. *International Review of Administrative Sciences*, 85(4):763–779. SAGE Publications. https://doi.org/10.1177/0020852317747370

Laux, J., Wachter, S., and Mittelstadt, B. (2023). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. *Regulation & Governance*, 18(1):3–32. John Wiley & Sons. https://doi.org/10.1111/rego.12512

Liboiron, M. (2021). *Pollution is colonialism*. Duke University Press. https://www.dukeupress.edu/pollution-is-colonialism

Lopes, I. M., Guarda, T., and Oliveira, P. (2019). How ISO 27001 Can Help Achieve GDPR Compliance. In *2019 14th Iberian Conference on Information Systems and Technologies (CISTI)*, pages 1–6. https://doi.org/10.23919/CISTI.2019.8760937

Marchant, G. E. (2011). Addressing the Pacing Problem. In Marchant, G. E., Allenby, B. R., and Herkert, J. R., editors, *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem*, volume 7, pages 199–205. Springer Netherlands, springer edition. https://doi.org/10.1007/978-94-007-1356-7_13

Maxwell, W. (2022). Les modes de régulation des activités numériques : exploration des tensions entre l'approche par les risques (risk-based) et l'approche fondée sur la protection des droits. Université Paris 1 Panthéon- Sorbonne. https://hal.science/tel-04026744

Mazzini, G. and Scalzo, S. (2023). The Proposal for the Artificial Intelligence Act: Considerations around Some Key Concepts. In Camardi, C., editor, *La via europea per l'Intelligenza artificiale*. Cedam. https://dx.doi.org/10.2139/ssrn.4098809

Merriam-Webster Dictionary (n.d.). Autonomy. https://www.merriam-webster.com/dictionary/autonomy

OECD (2019). Recommendation of the Council on Artificial Intelligence. C/MIN(2019)3/FINAL. https://one.oecd.org/document/C/MIN(2019)3/FINAL/en/pdf

OECD (2024). Recommendation of the Council on Artificial Intelligence. C/MIN(2024)16/FINAL. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449

Orwat, C., Bareis, J., Folberth, A., Jahnel, J., and Wadephul, C. (2024). Normative Challenges of Risk Regulation of Artificial Intelligence. *Nanoethics*, 18(11). https://doi.org/10.1007/s11569-024-00454-9

Perez, C. C. (2020). *Invisible Women: Exposing Data Bias in a World Designed for Men*. Vintage Books, London. https://carolinecriadoperez.com/book/invisible-women/

Rawls, J. (1971). *A Theory of Justice*. Harvard University Press. https://doi.org/10.2307/j.ctvjf9z6v

Robotics Openletter (2017). Open letter to the European Commission – Artificial Intelligence and Robotics. https://robotics-openletter.eu/

Rodrigues, R., Barnard-Wills, D., Wright, D., De Hert, P., and Papakonstantinou, V. (2013). *EU privacy seals project: inventory and analysis of privacy certification schemes*. Publications Office of the European Union. JRC85092. https://dx.doi.org/10.2788/29861

Rothstein, H., Irving, P., Walden, T., and Yearsley, R. (2006). The risks of risk-based regulation: Insights from the environmental policy domain. *Environment International*, 32(8):1056–1065. https://doi.org/10.1016/j.envint.2006.06.008

Rott, P. (2019). Certification of Medical Devices: Lessons from the PIP Scandal. In Rott, P., editor, *Certification – Trust, Accountability, Liability*, volume 16, pages 189–211. Springer International Publishing, Studies in European Economic Law and Regulation. https://doi.org/10.1007/978-3-030-02499-4_9

Ruschemeier, H. (2023). AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. *ERA Forum*, 23(3):361–376. https://doi.org/10.1007/s12027-022-00725-6

Schuett, J. (2023a). Defining the scope of AI regulations. *Law, Innovation and Technology*, 15(1):60–82. https://doi.org/10.1080/17579961.2023.2184135

Schuett, J. (2023b). Risk Management in the Artificial Intelligence Act. *European Journal of Risk Regulation*, pages 1–19. Cambridge University Press. https://doi.org/10.1017/err.2023.1

Smuha, N. A. (2019). The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International*, 20(4):97–106. Verlag Dr. Otto Schmidt. https://doi.org/10.9785/cri-2019-200402

Smuha, N. A. (2021). Beyond a Human Rights-Based Approach to AI Governance: Promise, Pitfalls, Plea. *Philosophy & Technology*, 34(1):91–104. https://doi.org/10.1007/s13347-020-00403-w

Smuha, N. A., Ahmed-Rengers, E., Harkens, A., Li, W., MacLaren, J., Piselli, R., and Yeung, K. (2021). How the EU Can Achieve Legally Trustworthy AI: A Response to the European Commission's Proposal for an Artificial Intelligence Act. https://dx.doi.org/10.2139/ssrn.3899991

Smuha, N. A. and Yeung, K. (2024). The European Union's AI Act: beyond motherhood and apple pie? https://dx.doi.org/10.2139/ssrn.4874852

Vainionpää, F., Väyrynen, K., Lanamaki, A., and Bhandari, A. (2023). A Review of Challenges and Critiques of the European Artificial Intelligence Act (AIA). *International Conference on Information Systems (ICIS) 2023 Proceedings*. https://aisel.aisnet.org/icis2023/aiinbus/aiinbus/14

van Leeuwen, B. (2014). PIP Breast Implants, the EU's New Approach for Goods and Market Surveillance by Notified Bodies. *European Journal of Risk Regulation*, 5(3):338–350. JSTOR. https://www.jstor.org/stable/24323461

von der Leyen, U. (2019). *A Union that strives for more – My agenda for Europe – Political guidelines for the next European Commission 2019-2024*. Publications Office of the European Union. European Commission, Directorate-General for Communication. https://data.europa.eu/doi/10.2775/018127

Wachter, S. (2024). Limitations and Loopholes in the EU AI Act and AI Liability Directives: What This Means for the European Union, the United States, and Beyond. *Yale Journal of Law & Technology*, 26(3). https://yjolt.org/limitations-and-loopholes-eu-ai-act-and-ai-liability-directives-what-means-european-union-united

Wachter, S., Mittelstadt, B., and Russell, C. (2021). Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI. *Computer Law & Security Review*, 41(105567). https://doi.org/10.1016/j.clsr.2021.105567

Wentholt, I. M. E., Hoekstra, J. B. L., Zwart, A., and DeVries, J. H. (2005). Pendra goes Dutch: lessons for the CE mark in Europe. *Diabetologia*, 48(6):1055–1058. https://doi.org/10.1007/s00125-005-1754-y

Zenner, K., Marcus, J. S., and Sekut, K. (2024). A dataset on EU legislation for the digital world. Bruegel. https://www.bruegel.org/dataset/dataset-eu-legislation-digital-world