Reading the drafts of the AI Act with a technical lens

Anonymous Author(s)
Affiliation
Address
email

Abstract

The draft AI Act is an effort led by European institutions to regulate the deployment and use of artificial intelligence. It is a notably difficult task, in part due to the polysemy of concepts such as artificial intelligence, covering topics such as foundational models, optimisation routines and rule-based models, among others. Furthermore, it gives a prism by which we can observe the wide variety of stakes different actors are pushing for.

After an initial draft proposed by the Commission in 2021, the European Commission, Council and Parliament will now discuss and draft the final version as part of the *trilogue phase*. The existence of these three versions gives us a chance to understand the negociations happening between the different European institutions, and as such is an interesting look into the currents that shape the artificial intelligence ecosystem.

In this paper we focus on the Commission, Council and Parliament proposals for the Act, and read them with a technical lens. In particular, we examine the technical concepts mobilized in the Act, and contextualize them in the wider sociotechnical environment surrounding artificial intelligence. For each main concept, we make a comparative analysis of each version, highlighting their differences and their impact.

This paper is primarily geared towards computer scientists, data analysts and machine learning researchers, in order to clarify the tenets and decisions made in the current versions of the act.

1 Introduction

The Draft AI Act is a broad proposal to regulate artificial intelligence in the European Union, pursuing the increase in efforts of digital regulation under Ms. Von der Leyen's term. The process is accelerating now as both the European Council and the European Parliament have publicised their draft versions, effectively entering the trilogue phase, *i.e.* informal interinstitutional meetings that aim at producing a final version the Council and the Parliament can adopt.

As per this process, two texts have been produced from the European Commission draft (`COM/2021/206`), both resulting from readings of the same documents by two different institutions. These institutions each come with their sociolegal processes, habits and customs, and do not communicate with each other before the trilogue phase. This makes it a unique opportunity to examine multiple constructions over the same base draft.

From a research perspective, this gives the opportunity to study the Act *as it is written*, giving us a window to understand the positioning of the three main European institutions, all subject to different external and internal incentives: in the case of the Council, each member state decides on acceptable positions for each topic, for example; institutions such as the Parliament, being elected by European citizens, have a duty of representativity; etc. This comparative analysis of similar documents is

Submitted to 37th Conference on Neural Information Processing Systems (NeurIPS 2023). Do not distribute.

common among sociologists, especially in the fields of science and technology studies [1] and the study of controversies [2].

In the remainder, we will refer to the "Draft AI Act" to describe the general text, and will add the institution name each time we wish to discuss a specific version. For clarity, we will not use the official code names of the texts, such as `COM/2021/206` for the Commission version, and so on.

The goal of this paper is to provide an interpretive exegesis of the Draft AI acts with a technical point of view, and to contextualise the different actors' positions with respect to regulating artificial intelligence; our goal is to shed light for technical practitioners on the definitions introduced in the drafts and their technological framing. We intend this paper as a quick entry point into the Act and its context for machine learning practitioners.

We focus on the following research questions:

- **RQ1.** To what extent do the technical notions introduced in the Draft AI Act coincide with technical notions of artificial intelligence?
- **RQ2.** Which aspects are prioritized in the AI Act, and what arbitrations are made? How does it tie in into the broader context?

2 Related works

Since the Commission first drafted the AI Act in April 2021, several studies have attempted to analyze its ins and outs. [3] first analyzed the risk-based approach of the AI Act and what different systems fall into which category, and with which requirements. [4] criticizes the lack of regulations for low-risk and zero-risk systems, which leaves the field open to loopholes. He also warns against overly idealistic requirements, such as the completeness and correctness of data sets. Similarly, [5] warns against defining high-risk AI too narrowly, leaving out systems that we would need to worry about. This notion of high-risk AI system, which is at the heart of the AI Act, was further examined by [6], who looked at the specific criteria that must be met to be considered high-risk. [7] surveyed startups to find out the impact of the AI Act. They found that around a third could be considered to be making high-risk AI systems, and almost half would be making general-purpose AI systems. Finally, [8] notes that the Act's risk-based approach, in particular the definition of high-risk applications by means of a list, is too arbitrary.

The reliance of the AI Act on technical standards was also scrutinized. [9] examines the AI Act's compliance regime and the predominance of delegation to standards for all tech-related topics. [10] is concerned by the absence of standards for AI today, given the important role they should play in defining the technical aspects of the Act.

Closer to our work, articles have attempted to analyze the definitions of AI under the Commission's proposal. [11] analyzes the difficulties of defining AI, including the existence of different definitions in different disciplines, the variety of AI types, and the legal stakes of defining AI in law. They further look at the definition of AI in the Commission version of the AI Act and what it cover, concluding that the definition is very broad, encompassing many types of computer programs. [12] compares possible definitions and classifications of AI systems in various texts by various actors, such as the Commission in the AI Act, the IEEE organization, AlgorithmWatch, the DEK, the OECD, and so on. [13] also lists popular definitions of AI proposed by computer scientists and philosophers, and how the definition proposed by the AI Act relates to them.

Some actors do not even think that AI should be defined in the Act. ETSI called the European Commission to leave the definition of AI to technical standards [14]. [13] also argues that the definition of AI in the AI Act does not meet the requirements for legal definitions, such as inclusiveness or precision. Indeed, he believes the definition is too broad, including systems with very different risks profiles that should not be treated the same way. He then proposes to push the risk based approach further, by not even defining AI but rather the risks that needs to be reduced.

Most of these studies focus on the Commission version of the AI Act but some dare to look at the other two versions. [15] discusses the Parliament's version, and in particular the new concept on foundational models. [16] looks at potential disagreements between EU institutions, pointing out notably the use of biometric surveillance in public spaces, the definition of high risk AI and of generative AI, and the conditions of enforcement of the text. However, to the best of our knowledge,

the strict comparison of all the definitions in the drafts has never been carried out. To do this, we introduce the different concepts with a technical reading.

Furthermore, using graphs as a formal model for texts (legal or not) is a common way to understand their structure, their relevant or redundant parts, and so on, including on legal texts [17]. This is typically done by extracting some notion of *edge* from the text, be it references from academic articles [18], legal texts [17], Wikipedia pages [19], among many others.

3 Legislative process in the European Union

Let us briefly describe the legislative process followed in the European Union, as it applies to the AI Act. Following meetings by large committees of experts from multiple horizons, as well as internal experts, the first draft of the text is written by the European Commission, and is made public. From this basis, both the Council and the Parliament write their own versions, adding, modifying or deleting elements from the Commission version. While multiple writers contribute to each draft, each institution appoints a *rédacteur général*, to harmonize the text and ensure its cohesiveness.

Once the three institutions have their version of the text, the *trilogue* phase begins: the three institutions discuss to draft the final text, until agreement. It is this final text that is to be [voted] by both the Parliament and the Council, and then be implemented in the Union.

In order to fuel the discussions, each version of the Act is to be translated in all of the Union's official languages. This is a particularly important step, as each text in each language has legal value.

4 Method

In order to identify the core definitions, we perform a systematic reading of the draft AI Acts, confronting articles in the Commission with their amendments by the Council and the Parliament. Notice that both the Council and Parliament preserve the numbering of the articles and recitals drafted by the Commission: the Parliament writes amendments, showing the original Commission text and its Parliament version, and the Council adds letters for insertions (*i.e.* an article added by the Council between Commission articles 3 and 4 will be labelled *4a*).

In order to identify key parts of the text, we model the document as a **citation graph** $G = (V, E)$, *i.e.* a set of nodes and edges between them. We build it as follows: each recital, article and title is a node, and we put a link between two nodes if they cite each other. For example, when the Draft Act states in Article 2(2): "For AI systems classified as high-risk AI systems in accordance with Articles 6(1) and 6(2) related to products covered by Union harmonisation legislation listed in Annex II, section B only Article 84 of this Regulation shall apply.", we add the following edges: (Article 2, Article 6), (Article 2, Annex II), (Article 2, Article 84). For any node $u \in V$, the *neighbourhood* of u is the set of nodes $\{v_0, v_1, \ldots, v_i\}$ (for $0 \leq i \leq |V|$) such that there is an edge between v and v_i in the graph. In formalism, for all $u \in V, N(u) = \{v : \exists (u, v) \in E\}$. The degree of u is defined as the number of neighbours of u, *i.e.* $d(u) = |N(u)|$. Finally, we say that node v is reachable from u if there exists a sequence of edges $((u, v_0), (u_1, v_1), \ldots, (u_i, v_i), (u_k, v))_{i=0}^{k}$ such that, for all $i \in [1, k], v_{i-1} = u_k$. For any given set of nodes $X \subseteq V$, we say that X is *connected component* if and only if for all nodes $u, v \in X$, v is reachable by u. Intuitively, a connected component corresponds to a distinct graph in a visualisation.

5 Defining artificial intelligence in the draft AI Acts

In this section, we focus on the main technical definitions that arise in the Draft Act. For each such definition, we check in which drafts it appears, and when a definition appears in multiple drafts, we outline the differences between each of them. Most of the definitions regarding artificial intelligence are in Article 3.

5.1 AI systems

We have here a prime example of the difficulty of defining artificial intelligence in precise terms; indeed, the Commission initially adopted a definition by example approach: AI systems are all types

of systems listed in the first Annex, classified in three categories: machine learning, logic-based systems, statistical learning. This circumvents the problem, as the Annex is easier to amend [1], and made it possible to list precise applications. Notice however that the systems so defined were still very broad. Take for instance objective function optimisation: one could argue that a sort function in a spreadsheet software fits the definition, even though most people would agree that it does not constitute *artificial intelligence*.

The Commission defines (Article 3, point 23) the notion of **substantial modification**: a (certified) AI systems needs a re-examination if such substantial modification happen. What makes a modification substantial is unclear from the draft (only defined as "a modification to the intended purpose for which the AI system has been assessed"), however both the Council and Parliament amend this definition by excluding modifications that have been planned in the initial assessment of the system.

In contrast, both the Council and the Parliament devise a more rigid in-text definition. They identify the presence of *elements of autonomy* as the key difference between AI and non-AI software. This likely encompasses supervised and unsupervised machine learning methods (that "autonomously" infer statistical biases from data) and reinforcement learning. Both Council and Parliament furthermore include "logic-based and symbolic methods".

The Council amends these definitions by notably adding the notion of **life cycle of an AI system**. It is defined as "the duration of an AI system, from design to retirement". This definition sets the scope of the Act with respect to AI systems, defining its framing [20].

5.2 Specific AI systems: general purpose and foundation models

Both the Council and the Parliament specify additional, more restrictive definitions of AI systems. In particular, the Council defines **general purpose AI systems** (GPAI as AI systems, open-source or not, that perform "generally applicable functions", such as image recognition, or speech processing; one immediately thinks of libraries such as `pytorch` or `scikit-learn` as GPAI. The Parliament defines **foundation models** as "an AI system model that is *trained on broad data at scale*, is designed for *generality of output*, and can be adapted to a *wide range of distinctive tasks*;" (emphasis ours). This definition appears less precise and more specific than the one provided by the Council, and extremely oriented by the recent development in so-called "foundation models", a term which use is not widespread and has been coined unilaterally by the Stanford AI center (HCAI). Furthermore, the Parliament explicitly offers to exclude open-source systems from the scope of the Act; though the reasons are unkown to the authors, this could either be a question of feasability (since open-source software can be *a priori* modified by anyone), or a way of fostering the release of code to the public.

5.3 Prohibited and high-risk AI systems

Notably, Title II follows a risk-based approach to define **prohibited AI practices**, high-risk systems (subject to more control and regulation), the other systems being subject to less regulation. The rationale is exemplified by thinking of applications: it makes sense to regulate AI systems running on critical infrastructure, or dealing with protected personal data, in a different way than AI systems that perform more benign tasks such as playlist recommendation, for example.

The Commission prohibits practices that retort to subliminal techniques, as well as any practice that might cause harm to individuals, and real-time remote biometric identification systems. While the Council only marginally expands and edits the Commission's proposal, the Parliament increases the number of prohibited practices, offering to prohibit systems that evaluate the risk of natural persons of offending or reoffending in criminal activity (likely in response to the COMPAS system [21]), the creation or expansion of facial recognition databases from the internet or CCTV footage, emotion inference systems, or any prohibited practice in another EU law.

In order to define what constitutes **high-risk** AI systems, the three institutions agree on a list defined in an annex (Annex III), editable over time by the Commission; this is similar to the Commission's definition of AI systems, as elicited *supra*. The Parliament however amends the text by offering that the Commission, 6 months before the entry into force of the regulation, consult all reelvant stakeholders to identify high-risk systems.

[1] The Commission allows itself in Article 4 to "adopt delegated acts", within the conditions outlined in Article 73: the Commission grants itself power to amend the act indefinitely.

Commission	Council	Parliament
Software; defines a list of approaches in Annex I;	Systems with elements of autonomy;	Machine-based systems with varying levels of autonomy;
Given set of human objectives;	*infers* how to achieve a given set of objectives; machine learning and logic-based;	Explicit or implicit objectives;
	Machine or human-based inputs and data;	
Content generation, recommendations, predictions		Generates outputs: predictions, recommendations, decisions;
Influence on environments with which the system interacts;	Influence on environment	Influence on physical and virtual environments
	adds *general purpose AI systems*	adds *foundation models*

Table 1: Synthetic table of the definitions or artificial intelligence in the three proposals

Alongside these constraints, the Commission establishes **regulatory sandboxes** in Title V as a means of fostering innovation: these are spaces limited in time and under supervision rules so that AI systems' providers have a way of performing *real-life testing* of their systems. The conditions of such sandboxes are only loosely defined, leaving to each member state (and its appointed AI regulating authority) the task of defining such sandboxes.

Separates task from application

6 The structure of the draft AI Act

We show, in Figure 1, the citation graph built from the Council version. The graph makes the structure of the Draft AI Act extremely clear, as a risk-based approach to regulating artificial intelligence. Indeed, the most cited nodes (*i.e.* the nodes with the highest degree) are "Title 3, Chapter 2" (requirements for high-risk AI systems), Articles 43 (Conformity assessment), 4b (Requirements for general-purpose AI systems and obligations for providers of such systems) and 71 (Penalties), and Articles 63 and 65 (national implementations of the regulation for high-risk AI systems). The final goal of the AI Act being to outline the requirements needed to affix a *CE marking* to AI systems. This is, to the best of the authors' knowledge, the first time that a CE marking would be affixed to algorithmic systems. Notice that the text strictly separates the *task* from its *use*: tasks are defined, typically, as systems (Title I), while applications are covered only through the prism of prohibited and high-risk practices (Title II and Title III).

As a guide rule for reading, the three institutions define three poles of competence: the Commission first and foremost focuses on technical definitions, while the Council focuses on market conformity and the Parliament on the protection of fundamental rights.

7 The future AI Act in context

The AI Act is part of an already well-developed European legislative ecosystem. Other digital legislation include the General Data Protection Regulation (GDPR), adopted in 2016 to protect EU citizens' personal data, and the Data Markets Act and Data Services Act (DMA, DSA) adopted in 2022, to regulated large online platforms. The AI Act is also part of a "AI package" and will be published alongside other legislation, like the AI Liability Directive. While the AI Act sets ex ante requirements for AI systems before they can be distributed on the European market, the AI Liability Directive will guarantee ex post liability for all stakeholders and ensure that common product liability rules are adapted to these new technologies. Older legislation is also been updated to take account of AI systems, that is the case for the Machinery Regulation and the General Product Safety Regulation.

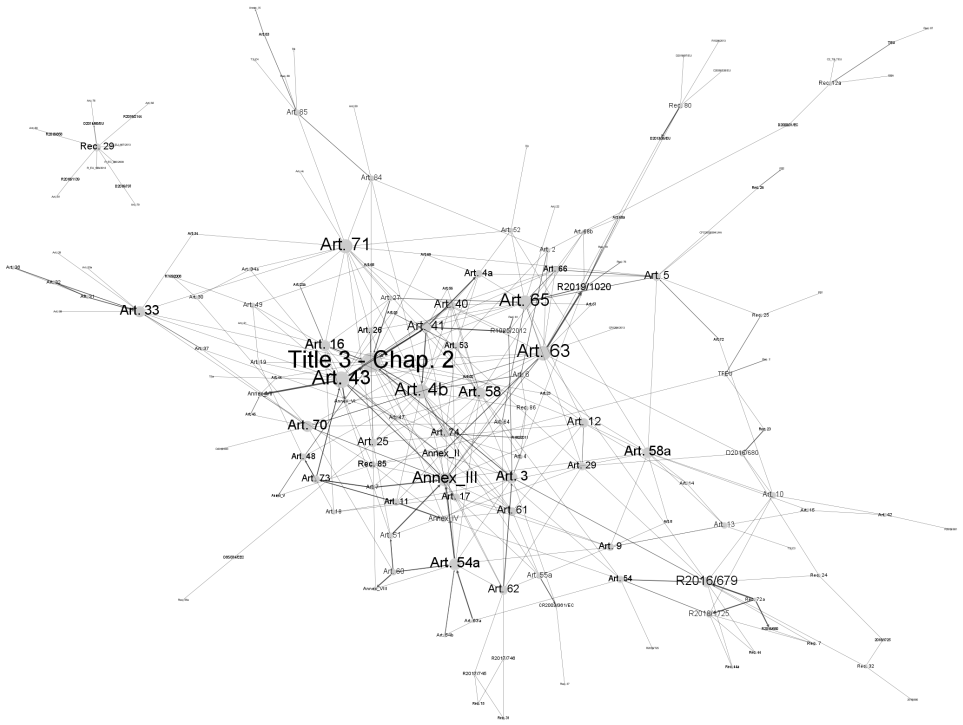

Figure 1: The graph extracted from the Council version of the Draft AI Act. Each node is an article, recital or part mentioned in the text, and there is a (directed) edge between two nodes u and v if article u cites article v. Node size is proportional to the number of times they have been cited (their *in-degree*). We restrict the graph to its two largest connected components.

Indeed, contrary to the GDPR, DSA or DMA, the AI Act will apply directly to products, i.e. items used by consumers. In this respect, the AI Act draws upon a set of legislation on product safety rules in Europe, known as the New Legislative Framework (NLF). Under the NLF, products must undergo a conformity assessment procedure, before been made available on the EU market. Conformity assessment results in the manufacturer signing a declaration of conformity and affixing the European Conformity (CE) mark to their product. This signifies that the product complies with all the requirements of European legislation for this type of product. To that end, manufacturers must refer to the technical specifications of their choices. The most commonly used technical specifications are harmonized standards, which are special types of European standards, drawn up by European Standardization Organizations (ESOs) following a standardization request from the European Commission.

Harmonized standards are intended to support EU legislation and to supplement legal requirements with technical means of compliance. Additionally, harmonized standards have a special status in European law, granting products that comply with them a "presumption of conformity" with the corresponding legislation and alleviating the burden of proof in the event of litigation. Although harmonized standards are technically voluntary, since manufacturers could choose other means of compliance, this advantage of presumption of conformity renders them almost mandatory for all economic players.

As part of the NLF, the AI Act also defines only the essential requirements to be met by any high-risk system. Manufacturers will have to choose technical specifications to comply with these requirements, and carry out the conformity assessment procedure. Even though harmonized standards do not exist yet for AI, ESOs are already working on them to complement the AI Act. The European Commission has publicly released on December 5, 2022 a draft standardization request, listing the topics to be addressed by ESOs in future harmonized standards, and corresponding to the obligations of high-risk AI systems set out in Title III, Chapter 2 of the AI Act.

8 Conclusion

In this paper, we focus on a discursive and comparative reading of the three versions of the draft AI Acts, as outlined by the European Commission, Council and Parliament. After recalling the process in which European laws are drafted and the relevant related work, we focus on the core technical definitions outlined in the drafts, exploring the commonalities and differences between the three versions of the text. We furthermore model the whole Act as a citation graph, highlighting the global structure of the Act. We finally replace the future Act in the broader context of digital regulations and standards put forth by the European Union.

We show both the general framing of the future AI Act – as a risk-based, market and compliance-oriented text –, and how each of the three institutions approaches the definitional challenge of artificial intelligence. This contributes to giving texture to the rich sociotechnical landscape of artificial intelligence, and to interfacing legal definitions with technical experts and practitioners, as a quick entry point into the Act and its context for machine learning practitioners.

Acknowledgments. The authors would like to thank [redacted for peer review] and [redacted for peer review], [redacted for peer review], as well as [redacted for peer review], for giving the authors of this paper the opportunity to discuss and present the basis of this work as [redacted for peer review].

References

[1] Sergio Sismondo. *An introduction to science and technology studies*, volume 1. Wiley-Blackwell Chichester, 2010.

[2] Michel Callon. Pour une sociologie des controverses technologiques. *Sociologie de la traduction: Textes fondateurs*, pages 135–157, 2013.

[3] Michael Veale and Frederik Zuiderveen Borgesius. Demystifying the draft eu artificial intelligence act—analysing the good, the bad, and the unclear elements of the proposed approach. *Computer Law Review International*, 22(4):97–112, 2021.

[4] Luciano Floridi. The European Legislation on AI: a Brief Analysis of its Philosophical Approach. *Philosophy & Technology*, 34(2):215–222, June 2021.

[5] Joanna J. Bryson. Europe Is in Danger of Using the Wrong Definition of AI. *Wired*, March 2022.

[6] Delaram Golpayegani, Harshvardhan J. Pandit, and Dave Lewis. To Be High-Risk, or Not To Be—Semantic Specifications and Implications of the AI Act's High-Risk AI Applications and Harmonised Standards. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '23, pages 905–915, New York, NY, USA, June 2023. Association for Computing Machinery.

[7] HubFranceIA. AI Act Impact Survey - Exploring the impact of the AI Act on Startups in Europe. Technical report, December 2022.

[8] Lilian Edwards. Expert opinion. Regulating AI in Europe: four problems and four solutions. Technical report, Ada Lovelace Institute, March 2022.

[9] Céline Castets-Renard and Philippe Besse. Ex ante Accountability of the AI Act: Between Certification and Standardization, in Pursuit of Fundamental Rights in the Country of Compliance. In Céline Castets-Renard and Jessica Eynard, editors, *Artificial Intelligence Law:*

Between Sectoral Rules and Comprehensive Regime. Comparative Law Perspectives. Bruylant, September 2022.

[10] Hadrien Pouget. The EU's AI Act Is Barreling Toward AI Standards That Do Not Exist. *Lawfare*, January 2023.

[11] Hannah Ruschemeier. AI as a challenge for legal regulation – the scope of application of the artificial intelligence act proposal. *ERA Forum*, 23(3):361–376, February 2023.

[12] Jakob Mökander, Margi Sheth, David S. Watson, and Luciano Floridi. The Switch, the Ladder, and the Matrix: Models for Classifying AI Systems. *Minds and Machines*, January 2023.

[13] Jonas Schuett. Defining the scope of AI regulations. *Law, Innovation and Technology*, 15(1):60–82, January 2023. Publisher: Routledge.

[14] Luca Bertuzzi. Standardisation body calls for AI definition, categorisation to be decided as standards, July 2022. Section: Technology.

[15] Meeri Haataja and Joanna J. Bryson. The European Parliament's AI Regulation: Should We Call It Progress? *Amicus Curiae*, 4(3):707–718, June 2023.

[16] Shana Lynch. Analyzing the European Union AI Act: What Works, What Needs Improvement, July 2023.

[17] Fabien Tarissan and Raphaëlle Nollez-Goldbach. The network of the international criminal court decisions as a complex system. In *Iscs 2013: Interdisciplinary symposium on complex systems*, pages 255–264. Springer, 2014.

[18] Ying Ding, Erjia Yan, Arthur Frazho, and James Caverlee. Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11):2229–2243, 2009.

[19] Tiphaine Viard, Thomas McLachlan, Hamidreza Ghader, and Satoshi Sekine. Classifying wikipedia in a fine-grained hierarchy: what graphs can contribute. *arXiv preprint arXiv:2001.07558*, 2020.

[20] Wanda J Orlikowski and Debra C Gash. Technological frames: making sense of information technology in organizations. *ACM Transactions on Information Systems (TOIS)*, 12(2):174–207, 1994.

[21] Tim Brennan, William Dieterich, and Beate Ehret. Evaluating the predictive validity of the compas risk and needs assessment system. *Criminal Justice and behavior*, 36(1):21–40, 2009.