The AI Act meets General Purpose AI: the good, the bad and the uncertain

Nídia Andrade Moreira[1]

s-njamoreira@ucp.pt

Pedro Miguel Freitas[1]

pfreitas@ucp.pt

Paulo Novais[2]

pjon@di.uminho.pt

[1] Universidade Católica Portuguesa, Faculty of Law, Católica Research Centre for the Future of the Law, Porto, Portugal
[2]Algorithm Centre/LASI, University of Minho, Braga, Portugal

Abstract. The general approach of the Draft of AI Act (December 2022) expanded the scope to explicitly include General Purpose Artificial Intelligence. This paper presents an overview of the new proposals and analyzes their implications. Although the proposed regulation has the merit of regulating an expanding field that can be applied in different domains and on a large scale due to its dynamic context, it has some flaws. It is essential to ascertain whether we are dealing with a general-risk category or a specific category of high-risk. Moreover, we need to clarify the allocation of responsibilities and promote cooperation between different actors. Finally, exemptions to the regulation should be properly balanced to avoid liability gaps.

Keywords: AI Act, General Purpose, Artificial Intelligence.

1 AI Act: the regulation of GPAI

1.1 Context

In April 2021, the European Commission published the Draft Proposal of AI Act (AIA), aimed at establishing a coordinated European approach to addressing the human and ethical implications of AI.

Throughout the law-making process[1], governments, experts and stakeholders formulated amendment proposals to improve it, highlighting the potential risks and misuse of technology and the need to protect innovation. With recent developments like GPT-4,

[1] The procedure 2021/016(COD) can be followed at https://eur-lex.europa.eu/procedure/EN/2021_106.

the so-called “General Purpose Artificial Intelligence” (GPAI) has become a topic of discussion.

The initial Draft (April 2021) did not explicitly reference GPAI, but this did not necessarily exclude it from the scope of the AIA. If a GPAI system entailed a high-risk purpose, it had to comply with the requirements of high-risk AI.

During the Slovenian Presidency (November 2021), a proposal was made to exclude an automatic application of the AIA to the development and use of GPAI. The AIA would only apply if the GPAI system had an intended purpose within the meaning of the AIA or if it was integrated into an AI system that was subjected to the AIA (article 52a and recital 70a).[2] However, the French Presidency (May 2022) proposed to expand the scope of the AIA to include these systems and explicitly regulate them to promote the safe development of AI.[3]. The new proposal adapted the requirements of high-risk AI to GPAI systems, unless the provider of the GPAI excluded any high-risk uses in the documentation accompanying the GPAI.

The French proposal was a major development but, in the meantime, the Czech Presidency made some amendments to the AIA proposal. Specifically, the direct application of the requirements for high-risk AI systems was replaced with the possibility of future implementing acts. Discussions are currently underway based on the General Approach (December 2022) prepared by the Czech Presidency.

In this paper, we analyze how the General approach (December 2022) regulates such systems.

1.2 Definition: dimensions of generality

AI models typically exhibit narrow capabilities and are designed or trained for specific tasks (fixed-purpose systems). However, we are now witnessing the emergence of AI systems that lack an intended and specific purpose and can be adapted to different tasks and contexts. These systems can perform tasks that were not foreseen by their creators.

Considering this reality, the latest versions of the AIA proposal explicitly mention “generative AI systems” (article 3(1)), but more importantly, they introduced the concept of “general purpose AI system (article 3 (1b)). The definition of a GPAI refers to an AI system that is intended to “perform generally applicable functions” and may be “used in a plurality of contexts” and be “integrated in plurality of other AI systems”. Therefore, the key elements of the GPAI definition seem to include (i) a range of purposes, (ii) the ability to operate in various contexts and (iii) integration into other AI systems, namely high-risk systems. However, are these definitional elements unique to GPAI and are they be cumulative?

---

[2] Presidency compromise text (Brussels, 29 November 2021) - see recital 70a and title IVA. According to ALLAI [1, pp. 13-14] the Presidency considers that it is impossible for providers to comply with requirements of high-risk because GPAI does not have a “intended purpose” (article 8 (2)). However, as mentioned by ALLAI, it is possible to reasonably foresee its use.

[3] 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 (13 May 2022). The final compromise text of French Presidency was released in 15 June 2022.

The concept of generality in the context of AI itself is complex.[4] Some AI systems, like Stable Diffusion or Midjourney have different and combined *abilities* in image generation (text2img, outpaiting, inpaiting, img2img, img2text). Other AI systems like GPT-4 possess specific ability such as natural language processing and perform a wide range of *tasks* [24], e.g. creating social media content, summarizing text, translating text, writing code, and more recently, accepting images as prompts; and can also be employed in various *domains* such as education or law[5].

The proposed definition is worded in a way that is not sufficiently useful and clear. It remains unclear whether the requisites should be seen as cumulative for an AI system to be classified as a GPAI, or if the characteristic of generality should be dependent on the intentionality behind the AI system.

Recital 12c states that GPAI "are AI systems that are intended by the provider to perform generally applicable functions, such as image/speech recognition, and in a plurality of contexts". Although this definition is clearer in the sense that a GPAI implies the provider's intentionality towards a variety of purposes and contexts, it is not identical to the one found in article 3 (1b), in several ways.

The particularity of GPAI systems is that they can be used for different tasks (multi-purpose systems) in different domains and with different types of input [15, p. 3]. Furthermore, these models can be seen as "Foundation Models" [4], serving as a base for downstream applications or tasks.[6]

According to this, a task-approach[17], along with a taxonomy-based approach should be taken. Even when trained for a specific task, GPAI could perform a wide variety of tasks, some of which were not even intended from the start. The crucial point is that an AI system can serve various purposes depending on the context and can be integrated into different systems. This is possible due to specific characteristics that make them powerful and flexible models, such as their large scale and abstraction capabilities.

Despite their differences from traditional systems, as they can perform a variety of tasks with minimal fine-tuning, the versatility of these models should not be confused with Artificial General Intelligence (AGI).[7] These models "are unable to generalize to completely different data types outside of their training data" [12, p.17]. Therefore, even though we are witnessing a significant evolution of technology, it is not, however, a major breakthrough.

These AI systems are large-scale models trained on extensive datasets, capable of accepting different types of inputs and possess a high number of parameters. They can

[4] Gutierrez et al. [17, p. 2] refers to four alternatives: ability, domain, task and output considering that is the task the key to define GPAI.

[5] For example, the genesis.studio developed the GPJ (Practical Guide to Justice) with the support of Microsoft Portugal to be used by the Portuguese Ministry of Justice. Is a chatbot platform based on ChatGPT that answers legal questions in natural language and it aims to explain how to initiate divorce proceeding and bridge the gap between the justice system and citizens [16].

[6] See the definition of [1, p. 12]. Recently a new amendment adopted by the European Parliament on 14 June 2023 introduces specific obligations for providers of foundation models – article 28 b.

[7] Madiega [25] considers that they are part of a new wave of AGI technologies.

also be fine-tuned to perform multiple tasks.[8] They serve as a pre-products that can be tailored to specific purposes or act as adaptable systems, serving as a base model that can be adjusted for different tasks.[9]

The definition is provided by the AIA is overly inclusive [18, p.4][10]. Our suggested approach aims to refine the proposed definition by excluding systems that were specifically designed and trained for tasks like speech recognition but can be (i) used for different purposes in various of contexts and (ii) integrated into other AI systems. These systems should not be classified as GPAI since they have a different model structure. The use of the system in different domains or within different AI systems should be a necessary condition, but not sufficient to classify it as a GPAI [18, p. 5]. However, even in this case, we may question how many tasks are necessary to classify an AI system as a GPAI. Should we adopt a quantitative perspective that considers to the level of capabilities or accuracy for tasks [18, p.5] or should we employ other criteria, such as a taxonomy-based approach? The introduction of specific rules for GPAI requires a clear definition as a starting point.

1.3 Regulation: challenges and risks

Despite the numerous benefits of these models[11], there are potential risks[12] and difficult choices regarding their regulation [25]. Since these models are trained on large datasets, often referred to as big data[9], they face specific challenges related to quality[13] and security throughout the data lifecycle [26, p. 1600]. As Foundation Models, any flaws in the base model and in data governance can have implications for later applications. Consequently, these AI systems can amplify biases and discrimination[14] found in the training data. Therefore, it is crucial to adopt data governance practices that require the use of curation techniques to measure bias[15], filter and label the data[16], ensuring that the final model meets quality requirements.[17]

Other risks must be addressed, such as copyright infringement [5], the generation of

---

[8] One popular example is GPT-3 that was trained to predict the next word of a sentence and then has adapted to answer question, translate and other tasks. Launched in March 2023, GPT-4 is more powerful and performs even more complex tasks [30].

[9] See [6]. LLMs can be important in the development of general language systems.

[10] The definition has been shortened in the recent amendment proposed by the European Parliament on 14 June 2023 - see Article 3(1d).

[11] For example, an experimental study has shown that ChatGPT can increase productivity and equality between workers [28].

[12] See [40]. Analyzing the risks of GPT-4, see [29].

[13] For this reason, some authors proposed a quality assessment method that attends to big data's characteristics. See the proposal of [8].

[14] See examples in [3].

[15] See [27].

[16] It could include pre-moderation or other techniques that filter data, detect and remove some content. This takes us to another ethical problem associated to the creation of AI related to workers exposed to such sensitive contact – as hate speech, images of sexual violence – for which they are not given extra care and are poorly paid [32].

[17] Some bias can be explained not only by the number of inputs or the quality of data but also by the way data is labeled or trained [41].

disinformation at scale [7], criminal misuse [14] and other potential damages that are difficult to enumerate[18], such as encouraging physical harm or what Kolt [20] coined as "black swans" damages, which are highly consequential risks that are challenging to predict in advance but easy to explain in hindsight.

Privacy concerns should also receive specific attention. Since GPAI models are trained on scraped data from the internet, issues related to transparency and consent arise in data collection and processing.

However, the absence of regulation of GPAI could stifle innovation and competition. As mentioned earlier, these systems can be adapted for downstream tasks. Without regulation, the responsibility of complying with the AI Act falls on the downstream users, which could be "too much of a burden, especially for SME's and micro enterprises" or perhaps even technically impossible [1, p. 14]. In such scenario, the market would be (further) dominated by big tech companies.

Considering that the original developers often possess greater resources and knowledge compared to downstream providers, it is essential to rethink the value chain, responsibilities and cooperation. Otherwise, the non-regulation of GPAI would exempt the creators of the GPAI from responsibility and shift the focus solely to downstream applications [12, p. 23].[19]

2 AIA Draft

2.1 AI requirements and obligations

In order to address the specific characteristics of these systems, a new title – "General Purpose AI Systems" – has been added to the AIA draft, which establishes specific requirements for GPAI systems (article 4a – article 4c).

While there are some other articles in the AIA draft that are directly applicable to GPAI systems, such as articles 5, 52, 53 and 69, the core requirements and obligations that these systems must comply with are defined in article 4b. It is presumed that GPAI systems have a high-risk use if they possess such capability unless the provider explicitly excludes all high-risk uses in the instructions or information related to the GPI system.

Therefore, unless the GPAI system is prohibited (article 5) or cannot be used in a high-risk manner or as a component of a high-risk AI system, or if its high-risk uses have been excluded by the provider, it must meet certain requirements (title III, chapter 2) which will be described in a future implementing act (article 4b (1)).

AIA draft adopts a risk-based approach, which classifies risk based on the "intended purpose" for which the GPAI system was developed. However, this type of approach presents certain difficulties.

One issue is that by focusing on regulating specific uses of AI and disregarding the underlying foundation models, a loophole is created for GPAI systems [34, p. 369].

[18] Other risks come from a climate policy perspective [39].

[19] Engler and Renda [12] consider that the division of responsibility should be based on the "cheapest cost avoider". This means that we should analyze which entity is best positioned to identify and mitigate risks at the moment they are most easily identifiable.

There can also be a discrepancy between the indicated purpose of an AI system and its actual purpose, as it may be used for different purposes that originally intended. A key characteristic of GPAI models is precisely their lack of a specific intended purpose. Consequently, the risk classification should consider "foreseeable purpose(s)" [13, p. 3, 33, p. 67]. In such cases, the provider of a GPAI should explicitly state those purposes during the conformity assessment.[20] Finally, if a GPAI is designed to perform multiple tasks without a specific intended purpose, it could potentially be used for any high-risk application.[21,]

As it stands, two options become evident: (i) identify and map all foreseeable uses and only apply the requirements if a high-risk purpose is identified, or (ii) consider that a high-risk purpose could be implicit, leading to the application of requirements to all GPAI systems.

In practice, it is likely that users, rather than providers, will determine the uses of GPAIs. Consequently, it appears that all GPAIs could fall under this category since, as base models, they can be utilized as high-risk AI systems or a components of AI high-risk system. If a GPAI lacks an intended purpose and the provider does not exclude a high-risk usage, then it *may be* employed in high-risk applications. Otherwise, we would be left with an ineffective risk-based approach.

Regarding "Foundational Models", one possible solution is to establish limits from the beginning, considering that they could be utilized for such purposes at any time. However, this approach risks over-regulating GPAI models [19, p.3], as it would mean applying specific requirements to all GPAI models. To address this, Helberger and Diakopoulos [19] propose a new approach to regulating these models: a general-risk category. This may have been the intention behind the wording of a specific title for the regulation of these models, but its interpretation remains unclear [2].

As previously mentioned, GPAI systems are required to fulfill certain requirements (article 4b (1) and impose specific obligations on providers (article 4b (2-6), although not all of these requirements are clearly defined.

For instance, when examining the obligations outlined in Chapter 2, Title II, which will be further specified in future implementing acts based on the "characteristics, technical feasibility, specificities of the AI value chain and market and technological developments", there is a possibility of unforeseen risks that have not been taken into account, often referred to as "black swan" risks.

Given the potential risks associated with GPAI systems, particularly those with significant economic and social impact, it is important to approach the establishment of specific requirements for these systems with caution. The purpose of these requirements should be to act as preventive measures against risks, which should be assessed based on both the potential uses of such systems and the possibilities of misuse. This depends on the fulfilment of a prior obligation of analysis of the misuses. However, it is important to note that we can only reasonably foresee certain risks [15, p. 6] and not all potential risks.

Requirements such as the risk management (article 4b (6), article 9(2)) and ensuring performance, robustness, and cybersecurity (article 15(1)) of these models should be

[20] Another problem is who, and when, should label an AI system as high-risk [33, p. 67-68].

[21] Engler and Renda [12, p. 20-21] consider that all GPAI systems would trigger these requirements.

based on reasonably foreseeable risks rather than being overly stringent, as exhaustively addressing all risks may be unviable or impossible. It may be preferable to identify specific sensitive scenarios that require scrutiny. However, certain requirements will likely necessitate ongoing monitoring to keep up with evolving developments.

Regarding conformity assessment procedures before deploying GPAI systems in the market or putting them into service (article 4b (2-3), article 16 (e)), multi-purpose systems raise some questions. Conducting separate conformity assessments for each possible use could be expensive or even impossible, especially if the creator is unaware of all downstream uses. One possible solution could involve imposing a duty on providers to indicate the foreseen uses of the model at the time of its creation and to distinguish between safe and unsafe uses.[22] Based on this information, providers could then recommend measures to mitigate risks associated with downstream applications in domains they consider safe.

Lastly, it is worth noting that article 4b (2) does not explicitly mention all the obligations outlined in article 16, omitting obligations such as keeping automatically generated logs by an AI (article 16d) or informing the relevant national competent authority in case of adopting corrective actions (article 16h). The reason for these omissions while including other obligations remains unclear.

2.2 Key elements: value chain and cooperation

One of the issues with the draft regulation is the uncertainty surrounding the responsibility for complying with the requirements of a GPAI system. If a provider makes significant modifications to a high-risk system, they should be subject to obligations (article 3(23), article 23a (1c)) while the upstream provider would not have these obligations (article 23 a (3).

According to Engler and Renda [12, p. 18], GPAI models typically require retraining and fine-tuning, which qualifies as a substantial modification. This leads to differing responsibilities. Kolt [20, p.33] expresses concern about the allocation of responsibilities in the new Draft, as it assigns responsibilities to entities with fewer resources and ability to mitigate risks, while exempting organizations with greater resources and expertise – the creators. Big tech suppliers play a crucial role in this context, as they possess technical control and better resources to understand, modify and test the models [10, p.10].

Policymakers should differentiate between the various entities involved in order to establish different obligations. Hacker et al. [18] suggests that four entities - developers (providers), deployers (users or providers), users (professional or non-professional) and recipients– should have different responsibilities. Some requirements apply universally to all AI systems and must be met from the beginning of the lifecycle, while others depend on the specific use of AI. In the former case, the requirements should be imposed on developers – *e.g.*, developers should comply with non-discrimination laws and data governance. In the latter cases, the requirements should be tailored to those who deploy and use such models – e.*g*. risk management should be the responsibility of deployers, taking into account the specific use.

[22] Similar to [12, p. 27]. They propose the adoption of a code of conduct created by the initial providers.

The contractual relationship between the provider and user is a crucial factor in allocating responsibilities. However, it also has the potential to create power imbalances. To promote fairness and cooperation among actors, Helberger and Diakopoulos [19, p. 5] suggest adopting mechanisms for regulatory scrutiny of contractual terms[23]. While this may be a suitable solution, there should be a greater balance in the distribution of obligations, fostering an environment of cooperation throughout the lifecycle of these systems. Users also have obligations in this regard and must publicly declare their use of such tools for professional purposes.

Understanding the lifecycle of these systems is essential in clarifying the responsibilities of the different actors based on the various components of the process, the actors involved in the value chain, the level of control (including downstream provider access)[24] and technological capabilities to "ensure a fair sharing of responsibilities along the AI value chain" (recital 12c). However, the new regulation may potentially exempt everyone from responsibility, as it excludes this obligation for small and medium-sized enterprises (article 55a (3)).

Companies with advanced technology can easily make technological adjustments to comply with AIA requirements and, in fact, some requirements – e.g., transparency – need to be observed from the early stages of development [18, p. 9].

The level of technical knowledge about the model will also be relevant: It is different (i) when downstream providers have limited access to the AI system without knowledge of the technical details, compared to (ii) having total access to the model and technical documentation. In the former case, continuous cooperation between the provider and downstream developer is required, expanding the responsibility of the upstream provider to control access and prevent misuse. In the latter case, the prevention of misuse should be addressed through contractual stipulations in advance.

This could lead us to another discussion about the release and research access of the models that has been ongoing[25]. If the models were not released, the risk would be better contained, but with public access, there is a better understanding of the risks, while it allows for algorithms' auditing by third parties [22]. Structured access [36] or the existence of a review board ([22] are solutions that deserve some thought.

Regulators also need to give special thought to open-source models. It is true that public access can improve innovation and promote cross-examination of the source code which can be valuable towards ethical AI [11][26], but on the other hand, it makes risk control incredibly difficult [37, p. 4] and creates more opportunities for malicious uses and cyberattacks.

The responsibility of providers could be different according to the release procedure and option for the AI. Limited access potentially blocks or diminishes high-risk or out-of-scope uses, although there are loopholes, as mentioned by Solaiman [37, p.5], because users can share access with unauthorized users. These can be seen a kind of know-your-customer approach. In this case, the developers can limit who will use their

[23] That seems to be the intention with the proposed article 28a of the draft regulation adopted by the European Parliament on 14 June 2023.

[24] This is more complex because there will be different relations between providers and different levels of control of the system according to the strategy adopted to distribute the GPAI [21].

[25] For example, see [31, 35, 38]).

[26] Otherwise, we could have a concentrated power of organizations [37, p. 3].

models and define how they can be used, confirming this periodically.

Considering the downstream use of the GPAI on high-risk systems, the *cooperation* between providers was considered a key element of regulation. Article 4b (5) determine that providers should provide "necessary information to other providers intending to put into service or place such systems on the Union market as high-risk AI systems or as a component of high-risk AI systems" for the latter to meet the requirements.

This legal basis of cooperation encourages cooperation between providers, especially if a GPAI will be used for a specific high-risk purpose. The Commission should adopt an implementing Act to define what is "necessary information". At least, the provider should be obligated to provide information – but also instruction – about the safety of the system.

This cooperation is essential because we are faced with a paradigm shift in relations among providers and users that is characterized by the *interdependence* between upstream and downstream providers.

Trade secrets or IP rights must be protected, which has led some authors to propose the adoption of protective measures such as nondisclosure agreements or access to the information under certain conditions [18, p. 10]. Protective measures are essential to encourage cooperation.

However, this obligation occurs only at an initial stage – "to put into service or place such systems on the Union market". Nevertheless, this obligation should not be limited to that moment but should also be extended to allow for continuous cooperation and monitoring of the system to mitigate its risks [12, p. 26]. Therefore, this obligation must be complemented by the requirement for periodic and regular mandatory assessment on the risks of the system, including potential new uses.

2.3 Exemptions

The obligations established on article 4b are not applied if the provider "has explicitly **excluded** all high-risk uses in the instructions of use or information accompanying" (article 4c (1)). However, if providers consider that there may be misuse, the requirements established on article 4b apply (article 4c (2)). If the providers are aware – whether detected or informed – of any misuse, they should adopt measures to prevent further misuse (article 4c (3)).

It is difficult for providers to rely on the exception stated on article 4c (1), at least in good faith. Unless a provider is technically able to exclude high-risk use, the myriad of uses of a GPAI system means that it *may be* used for high-risk purposes [18, p. 5], and thus the obligations are applicable (article 4c (2).

The "notice-and-action mechanism" aligns with the post-monitoring of obligation (Title VIII, Chapter 1). Periodical risk monitoring assessments must be mandatory – similar to systemic risk monitoring approach outlined in article 34 of the Digital Services Act [19, p.4]. Additionally, other *ex post* measures should be considered, such as technical measures – *e.g.*, providers could disable access to certain users through APIs [12, p. 21] (although not applicable in the case of open-source).

Additionally, according to article 55a (3), these requirements and obligations do not apply to micro, small and medium-sized enterprises. However, this raises the question: if the objective of the AIA is to regulate technology that could have high impact, should SMEs be excluded from its scope? If SMEs wish to create GPAI systems, shouldn't

they follow the same set of requirements and obligations, at least to some extent? The argument that compliance with the AIA's requirements and obligations might prove to be burdensome for SMEs, but a solution could be found either in government support to the companies with fewer resources or by establishing a simplified set of requirements and obligations.

3 Conclusions

GPAI in an expanding branch of AI. The development of GPAI systems is a prominent trend that needs to be regulated, taking into account their features and the complexity of the value chain. These increasingly powerful systems are multi-purpose and can be applied to tasks they were not initially trained for.

As mentioned, we can identify the potential risks and adopt some strategies to mitigate them. However, not all risks and harms can be anticipated from the beginning since it can be challenging or even impossible to predict all the uses. Therefore, the regulations should encompass the entire lifecycle of AI.

In case of a high-risk use, the requirements should align with those specified in Title III. The final model intended for high-risk purposes should have specific requirements regardless of the AI system used, but these requirements may need to be adapted through implementing acts. Other GPAI systems, that do not have a specific high-risk purpose but may adopt one, should have certain requirements and obligations based on their specific characteristics – a specific-risk approach.

However, the distinction between providers/developers and users/deployers does not neatly apply to these systems. In fact, there can be intermediate entities who adapt or fine-tune the model. As a result, the relationship between actors is more complex, as is the lifecycle of these systems.

Damages can arise during the entire lifecycle of AI, from the upstream development to concrete application. There has to be a balanced distribution of responsibilities, ensuring a fair and clear allocation. Some issues may arise from the very beginning, at the source of value chain. Therefore, at the very least, a fundamental rights assessment should be required, taking into consideration vulnerable groups, potential misuses and the need for upstream providers to address problems discovered downstream.

The AI Act may slow down the development of AI [23] but will not impede innovation. Instead of merely balancing innovation and trustworthiness or fearing new developments, the focus should be on the type of AI that we desire.

Regulation will govern the future of increasingly important systems. However, to avoid stifling innovation and to encourage cooperation, regulations should be proportionate and tailored to the nature of the sector and the stage of technological development. In this regard, regulators may seek the involvement of technology experts and companies operating in these emerging sectors to help develop more precise and suitable regulations. Therefore, a collaborative environment among different actors – companies, regulator, and other stakeholders – will foster responsible innovation and ensure a safe and ethical development of new technologies.

Acknowlegments

The work of Nídia Andrade Moreira has been supported by FCT - Fundação para a Ciência e Tecnologia within the Grant 2021.07986.BD and the Project UIDB/04859/2020. The work of Paulo Novais has been supported by FCT – Fundação para a Ciência e Tecnologia within the R&D Units Project Scope: UIDB/00319/2020.

References

1. ALLAI. AIA in-depth. Objective, Scope, Definition. Articles 1-4 & Annex I (2022).
2. Bertuzzi, L. AI Act: EU Parliament's crunch time on high-risl categorization, prohibited practices (2023). https://www.euractiv.com/section/artificial-intelligence/news/ai-act-eu-parliaments-crunch-time-on-high-risk-categorisation-prohibited-practices/, last accessed 2023/04/23.
3. Biddle, S. The Internet's New Favorite AI Proposes Torturing Iranians and Surveilling Mosques, The Intercept (2022). https://theintercept.com/2022/12/08/openai-chatgpt-ai-bias-ethics/, last accessed 2023/04/02.
4. Bommasani et al. On the Opportunities and Risks of Foundation Models (2022). https://arxiv.org/abs/2108.07258
5. Brittain, B. Lawsuits accuse AI content creators of misusing copyrighted work. Reuters (2023). https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/, last accessed 2023/04/02.
6. Brown, T., et al. A. Language models are few-shot learners. *Advances in neural information processing systems*, 33,1877- 1901 (2020).
7. Buchanan, B., et al. Lies and Automation. How Language Models Could Change Disinformation (2021).
8. Cai, L., Zhu, Y. The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. Data Science Journal, 14: 2, 1-10 (2015). http://dx.doi.org/10.5334/dsj-2015-002
9. Dhirani, L.L., et al. Ethical Dilemmas and Privacy Issues in Emerging Technologies: A Review. Sensors 23(3), 1151 (2023). https://doi.org/10.3390/s23031151
10. Edwards, L. Regulating AI in Europe: four problems and four solutions. Expert Opinion. Ada Lovelace Institute (2022).
11. Engler, A. How open-source software shapes AI policy. Report from The Brookings Institution's Artificial Intelligence and Emerging Technology (2021).
12. Engler, A.C., Renda, A. CEPS in-depth analysis. Reconciling the AI value chain with the EU's Artificial Intelligence Act (2022).
13. European Consumer Voice in Standardization (ANEC). ANEC comments the European Commission proposal for an Artificial Intelligence Act. Position Paper (2021).
14. Europol. ChatGPT. The impact of Large Language Models on Law Enforcement (2023)
15. Future of Life Institute. General Purpose AI and the AI Act (2022).
16. Genesis.studio. GPJ – A implementação do chatGPT para o Ministério da Justiça pelo genesis.stiduo (2023). https://genesis.studio/gpj-a-implementacao-do-chatgpt-para-o-ministerio-da-justica-pela-genesis-studio/, last accessed 2023/04/02.
17. Gutierrez, C.I., et al. A Proposal for a Definition of General Purpose Artificial Intelligence Systems (2022).
18. Hacker, P., Engel, A., Mauer, M. Regulating ChatGPT and other Large Generative AI Models. Working Paper (version April 5, 2023).Avaliable at https://arxiv.org/abs/2302.02337.
19. Helberger, N. and Diakopoulos, N. (2023). ChatGPT and the AI Act. *Internet Policy Review*, *12*(1).
20. Kolt, N., Algorithmic Black Swans. Washington University Law Review, Vol. 101, Forthcoming. (2023). Available at SSRN: https://ssrn.com/abstract=4370566.
21. Küspert , S., Moës, N., Dunlop , C. Ada Lovelace Institute Blog The value chain of general-

purpose AI (2023).
22. Liang, P.; Bommasani, R.; Creel, K.; Reich, R. The time is now to develop community norms for the release of foundation models. Stanford University Human-Centered Intelligence (2022).
23. Liebl, A., Klein, T. AI Act Impact Survey. Exploring the impact of the AI Act on Startups in Europe (2022).
24. Lim, R, Wu, M., Miller, L. Customizing GPT-3 for your application. OpenAI. (2021). https://openai.com/blog/customizing-gpt-3, last accessed 2023/04/02.
25. Madiega, T. General-purpose artificial intelligence. Digital issues in focus at a glance. European Parliamentary Research Service (2023).
26. Moura, J., Serrão, C. Security and Privacy Issues of Big Data. In I. Management Association (Ed.), Cloud Security: Concepts, Methodologies, Tools, and Applications, pp. 1598-1630 (2019) IGI Global. https://doi.org/10.4018/978-1-5225-8176-5.ch080
27. Nadeem, M., Bethke, A. and Reddy, S. StereoSet: Measuring stereotypical bias in pretrained language models. *arXiv preprint arXiv:2004.09456* (2020).
28. Noy, S. and Zhang, W., Experimental Evidence on the Productivity Effects of Generative Artificial Intelligence (2023).
29. OpenAI .GPT-4 System Card (2023). https://cdn.openai.com/papers/gpt-4-system-card.pdf
30. OpenAI . GPT-4 Technical Report (2023). https://arxiv.org/pdf/2303.08774.pdf
31. Ovadya, A.; Whittlestone, J. (Reducing malicious use of synthetic media research: Considerations and potential release practices for machine learning. arXiv preprint arXiv:1907.11274 (2019).
32. Perrigo, B. OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic, TIME (2023). https://time.com/6247678/openai-chatgpt-kenya-workers/ , last accessed 2023/04/02.
33. Raposo, V.L. The European Draft Regulation on Artificial Intelligence: Houston, We Have a Problem. In: Marreiros, G., Martins, et al.. (eds) Progress in Artificial Intelligence. EPIA 2022. Lecture Notes in Computer Science (), vol 13566. Springer, Cham (2022).
34. Ruschemeier. H. AI as a Challenge for legal regulation – the scope of application of the artificial intelligence act proposal. Era Forum. 23:361-376 (2023).
35. Sastry, G. Beyond "release" vs. "not release" (2021). https://crfm.stanford.edu/commentary/2021/10/18/sastry.html, last accessed 2023/04/03.
36. Shevlane, T. Structured Access: An Emerging Paradigm for Safe AI Deployment'. In Justin B. Bullock and others (eds), The Oxford Handbook of AI Governance (2022).
37. Solaiman, I. (The Gradient of Generative AI Release: Methods and Considerations (2023).
38. Staff, P. Managing the Risks of AI Research: Six Recommendations for Responsible Publication (2021).
39. Taddeo, M., Tsamados, A., Cowls, J. and Floridi, L. Artificial intelligence and the climate emergency: Opportunities, challenges, and recommendations. *One Earth*, 4, 6 (2021), 776-779.
40. Weidinger, L. et al., Ethical and social risks of harm from Language Models. (2021)
41. Zhao, Z., et al. *Calibrate before use: Improving few-shot performance of language models*. Proceedings of the 38th International Conference on Machine Learning, PMLR 139:12697-12706 (2021)