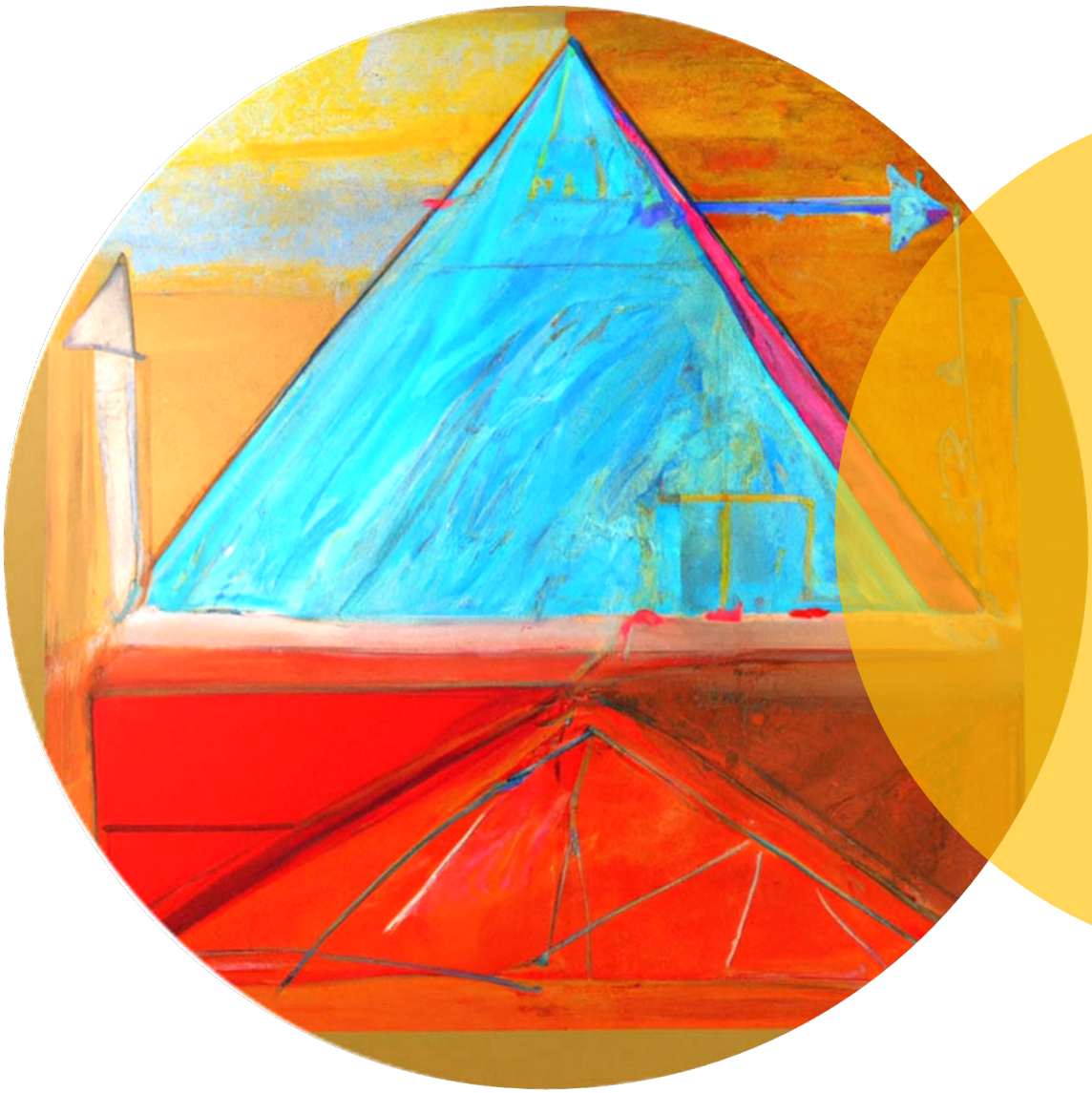
Tools for Navigating the EU AI Act: (2) Visualisation Pyramid

Theodore Christakis, Theodoros Karathanasis

AI-REGULATION.COM

Tools for Navigating the EU AI Act: (2) Visualisation Pyramid

By Theodore Christakis and Theodoros Karathanasis



Tools for Navigating the EU AI Act: (2) Visualisation Pyramid

The following article features a comprehensive visualization pyramid designed to illustrate the risk-based approach of the EU AI Act in a single, intuitive graphic. This tool is intended to be useful to academics, students, practitioners, data and AI enthusiasts, as well as anyone keenly interested in the imminent adoption of the EU AI Regulation.[1]

Following an extensive legislative process spanning nearly three years, the world's inaugural comprehensive regulatory framework for Artificial Intelligence – the European Union (EU) AI Regulation, commonly referred to as the "EU AI Act" – is on the brink of being adopted. The provisional agreement on the proposal, brokered in December between the Council presidency and negotiators from the European Parliament, was finalized and unanimously endorsed by the Council of the EU on February 2, 2024. The European Parliament's internal market and civil liberties committees gave their approval of the Council's finalized provisional agreement on February 13. The text awaits formal adoption at the Parliament's plenary session on March 13. Although it will come into force immediately after its publication in the EU Official Journal (expected in May), it will only become enforceable in two years, with the exception of certain provisions that will become applicable earlier.[1] Developers of AI applications, and companies and institutions engaged in AI must begin analyzing the text and commence preparation accordingly.

The significance of the AI Act cannot be overstated, as it represents a landmark regulatory milestone in the field of AI. Much like the General Data Protection Regulation (GDPR), which revolutionized data protection standards globally, the AI Act seeks to establish a comparable framework for AI technologies. With its ambitious scope and comprehensive regulations, the AI Act aims to ensure the responsible development, deployment and governance of AI systems in the EU. By setting clear rules and standards, the AI Act aims not only to protect fundamental rights and safety, but also to foster innovation and trust in AI technologies.

Moreover, the AI Act's ambition is to have a "Brussels effect"[2] similar to the one enjoyed by the GDPR. Just as the GDPR has set a benchmark for data protection worldwide, the AI Act's influence could extend far beyond Europe, shaping AI regulations and practices on a global scale.[3] This "Brussels effect" will underline the EU's leadership in setting the agenda for responsible AI development and governance, positioning it as a key player in shaping the future of AI regulation internationally.

Nevertheless, the AI Act is a large and complicated document, comprising, in its current non-definitive form, 252 pages, including numerous recitals, articles and annexes.[4] Its comprehensive nature underscores the complexity inherent in the regulation of AI technologies. Despite its thoroughness, many practitioners we have engaged with express a sense of being overwhelmed and perplexed by the intricacies of the EU AI Act. The volume of content, coupled with the complicated legal language, often leaves individuals feeling 'lost' amidst the nuances and technicalities of the Regulation. As such, navigating the AI Regulation is a significant challenge for practitioners seeking to understand and comply with its provisions effectively.

[1] In accordance with Article 85, the AI Regulation will enter into force twenty days after its publication in the Official Journal of the EU. It will become applicable 24 months later, with the exception of certain titles within the regulation that will become applicable in six or twelve months, based on Article 85(3).

[2] Bradford, A. (2020) 'The Brussels Effect: How the European Union Rules the World', Oxford University Press, 424 p.

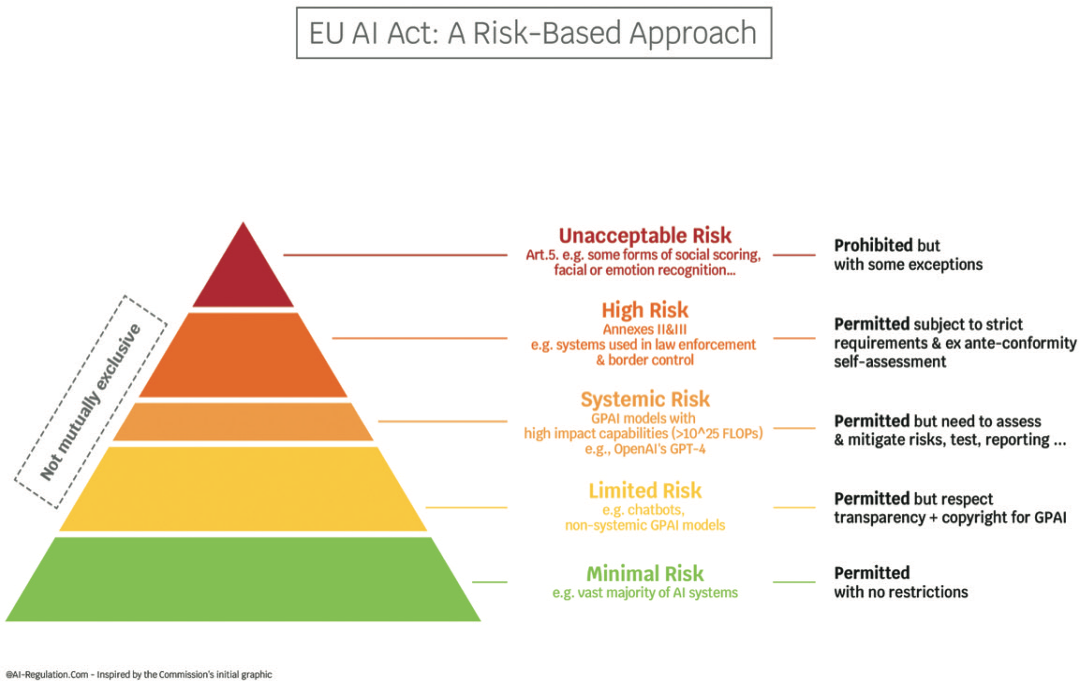
[3] See Almada M, Radu A. (2024) 'The Brussels Side-Effect: How the AI Act Can Reduce the Global Reach of EU Policy', German Law Journal, doi:10.1017/glj.2023.108; Siegmann, C, Anderljung M. (2022) 'The Brussels Effect and Artificial Intelligence: How EU regulation will impact the global AI market', Center for the Governance of AI, arXiv:2208.12645;

[4] The number of Recitals, Articles and Annexes mentioned in the present article are subject to modification upon final publication of the AI Act in the Official Journal (OJ) of the European Union.

Recognising this challenge, the AI Regulation Chair has taken the initiative of providing and developing tools and resources, such as the final AI Act along with a detailed and interactive Table of Contents (ToC), aimed at aiding practitioners navigate the complexities of the AI Act effectively.

These tools aim to provide clarity, guidance, and practical insights in order to assist stakeholders in understanding and complying with the regulation's provisions amidst its length and intricacy.

The following is a comprehensive visualisation pyramid designed to illuminate the intricate logic and core content of the EU AI Act in a single, intuitive graphic.



This paper is divided into three sections.

Part 1 elucidates the genesis of our pyramid, drawn from a presentation produced by the European Commission in 2021. This presentation underscored the Commission's commitment to a "risk-based" approach in its initial AI Act proposal, emphasising its intention not to impede AI advancement and innovation within the European Union. Our interpretation of the Commission's pyramid suggests that it integrates two crucial elements: one depicting criticality and the other depicting quantity and the breadth of AI applications subject to new EU regulations.

Part 2 discusses successively three important issues. *First*, it enquires whether and how the Commission's pyramid concept endured the legislators' incorporation of regulations for GPAI systems into the AI Act. This is not merely a question of illustration; it is a crucial examination of the survival of the underlying concept—the risk-based approach—upon which the AI Act was established. *Second*, having concluded that the Commission's pyramid is still standing, a new urgent question arises: where should one place "Systemic Risk", and more broadly, GPAI, within the Commission's pyramid? The article explains the reasons that pushed us to place "Systemic Risk" under the "High Risk" level. *Thirdly*, the article contends that while the foundational pyramid structure of the AI Regulation remains intact, significant expansion has occurred during the negotiation phase. The Commission's initial pyramid, which resembles the modest proportions of the pyramid of Menkaure, has thus evolved into a structure reminiscent of the pyramid of Khafre, if not the grandeur of the great pyramid of Khufu.

Part 3 provides a detailed examination of our visualisation pyramid, highlighting its principal enhancements and additions.

1) The Initial Pyramid Designed by the European Commission

The Commission designed in 2021 a four-layer pyramid graphic to illustrate the logic behind the proposed AI Regulation, emphasising its risk-based approach. While this pyramid has not been officially published by the Commission on its website, as far as we are aware, it has been utilized by members of the Commission in various workshops, such as here or here.[5] This visual representation aimed to elucidate the design of the regulatory framework, which emphasizes the assessment and management of risks associated with AI technologies.

Our interpretation of the Commission's pyramid suggests that it integrates two crucial elements: one depicting criticality and the other depicting quantity.

1.1. Criticality

Closely aligning with the German Data Ethics Commission's proposal, the European Commission's rationale involved determining the level of criticality of algorithmic systems through an overarching model.

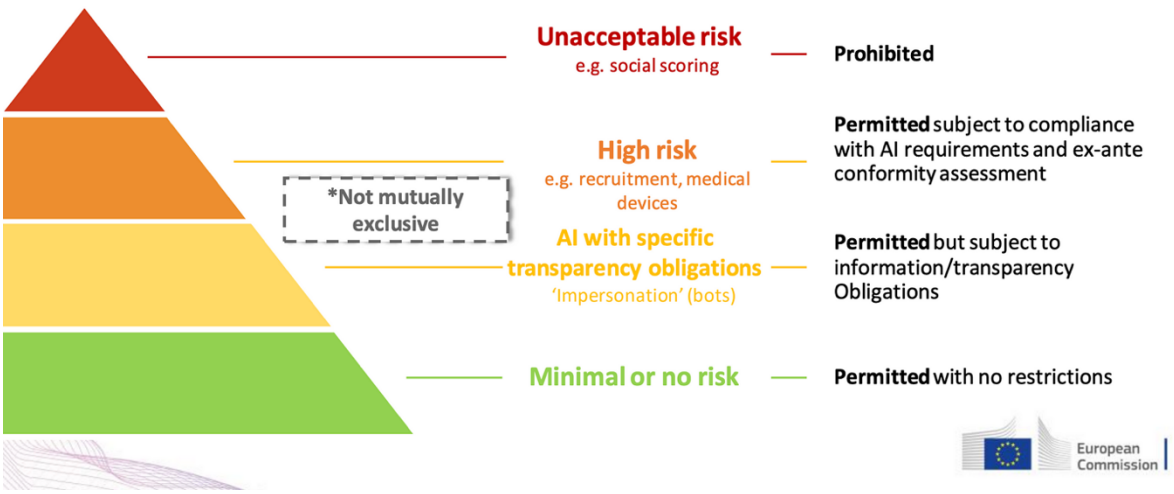
The European Commission's pyramid employs four distinct colors to delineate the varying levels of criticality for AI systems outlined in the initial AI Act proposal: red signifies 'Unacceptable Risk'; orange represents 'High Risk'; yellow denotes 'Low Risk'; and green indicates 'Minimal or No Risk.'

AI applications deemed to pose zero or negligible potential harm are situated in the green segment at the pyramid's base and are exempt from specific EU regulations. Progressing to the second tier, we encounter AI systems with limited potential harm, for which the Commission's proposal recommended only modest regulatory requirements, primarily focused on transparency obligations.

Ascending to the third level, we encounter AI applications with significant potential harm ('High Risk'), subject to a comprehensive set of articles in the Commission's proposal, introducing substantial requirements for developers and extensive oversight. At the pinnacle of the pyramid, denoted by red, lies the highest level of risk, *ipso facto* considered unacceptable and contrary to European values. This 'level 4' signifies AI applications that, according to the Commission's proposal, should be banned, with certain exceptions introduced.

While this interpretation of the pyramid primarily emphasizes criticality, we posit that the Commission also utilizes the pyramid structure to convey a message regarding the breadth of AI applications subject to new EU regulations. Thus, the pyramid serves not only to indicate criticality, as seen in the German Data Ethics Commission's model, but also to

The Commission's Initial Slide



[5] Before this, the German Data Ethics Commission had proposed, in December 2019, its own "criticality pyramid" for the regulation of algorithmic systems (see here, p. 177).

illustrate the distribution of AI applications based on their level of risk.

1.2. Quantity

As previously noted, the European Commission sought to convey the message that the newly proposed AI Regulation would achieve a delicate balance, protecting societal values while also fostering innovation. In response to concerns about the risks of overregulation,[6] which could impede the EU's competitiveness compared to leading AI nations like the US and China, the Commission aimed to dispel such criticisms. It wished to emphasize that the proposed AI Act would not hinder innovation; rather, it is designed to encourage innovation and technological advancement within the EU.[7] These regulatory measures, according to the logic of the Commission, are intended to minimize potential risks and establish a framework for trusted AI, thereby facilitating progress while ensuring the responsible development and deployment of artificial intelligence technologies.

The pyramid, then, wishes probably to serve a dual purpose beyond illustrating criticality; it also aims to depict the breadth of AI applications subject to new EU regulations.

At its base, depicted in green, lie the *majority* of AI applications posing minimal or no risk, thereby requiring no regulation. Moving upwards, a *significant* number of applications fall into the limited risk category, denoted in yellow, and are subject only to transparency requirements.

The Commission's pyramid communicates that the majority of AI applications are situated within the green or yellow layers, emphasizing that the development of AI in Europe will not be hindered by the AI Regulation. Rather, only a more limited number of riskier applications, subject to more stringent rules, are affected. At the pinnacle of the pyramid, depicted in red, reside a small number of AI applications deemed to pose unacceptable risks, such as real-time remote biometric identification in publicly accessible spaces by law enforcement agencies, which is prohibited except under specific exceptions. The second layer, highlighted in orange, encompasses high-risk applications, including a significant number of AI systems, albeit fewer than the green and yellow layers.

This pyramid made sense when it was designed in April 2021, soon after the publication of the AI Act draft by the European Commission. Its draft proposal, however, had not anticipated the commercialisation of powerful generative AI models - trained using a large amount of data and capable of performing a wide range of tasks (e.g., ChatGPT). The arrival of these generative AI models challenged the risk-based approach initially proposed by the Commission.

2) Conceptualising the Final AI Act: Is the Pyramid Still Standing?

The negotiating process surrounding the AI Act has been characterised by a number of significant events, each highlighting key aspects of the debate. But undoubtedly, the most significant event, was the commercialisation of Chat GPT in November 2022.

This was exactly around the time when the Council adopted its general approach in which it proposed adding a new, brief title to the Commission's draft to take account of situations where AI systems can be used for many different purposes (e.g., General Purpose AI - GPAI).[8]

[6] Madiega T. (2023) 'Artificial intelligence act', Briefing of the European Parliament Research Service, PE 698.792, June

[7] 'At the same time, the provisions of the regulation are not overly prescriptive and leave room for different levels of Member State action for elements that do not undermine the objectives of the initiative, in particular the internal organisation of the market surveillance system and the uptake of measures to foster innovation' (Explanatory Memorandum preceding the AI Act proposal, Section 2, point 2.4)

[8] The Council suggested introducing Article 4b(1) stipulating that certain requirements for high- risk AI systems would also apply to general purpose AI systems. However, the Council noted that "instead of direct application of these requirements, an implementing act would specify how they should be applied in relation to general purpose AI systems, based on a consultation and detailed impact assessment and taking into account specific characteristics of these systems and related value chain, technical feasibility and market and technological developments. The use of an implementing act will ensure that the Member States will be properly involved and will keep the final say on how the requirements will be applied in this context".

The introduction of Open AI's model - ChatGPT - to the market prompted the Parliament to recognise the urgent need for more substantive regulatory measures, leading to the inclusion of GPAIs within the AI Act as a response to evolving technological advancements and potential risks.[9]

The fundamental inquiry at hand is whether and how the Commission's pyramid concept endured the legislators' incorporation of regulations for GPAI systems into the AI Act. This is not merely a question of illustration; it is a crucial examination of the survival of the underlying concept—the risk-based approach—upon which the AI Act was established.

2.1. How the Pyramid Survived GPAI: The Invention of "Systemic Risk"

Addressing the challenge of integrating GPAI into the AI Act without undermining its core design as a risk pyramid based entirely on a risk-based approach, presented a complex dilemma and sparked heated discussions.

Initially, numerous proposals emerged to address this issue by categorizing GPAI as 'high risk' and placing it within the Commission's orange Level 3 of the pyramid. However, these proposals elicited considerable debate and opposition, with many arguing that such categorization was unwarranted.

Indeed, the underlying logic of the Commission's initial proposal focused on assessing the risk posed by a specific application based on its purpose, rather than on the AI system itself. It became evident that most applications of ChatGPT (or other GPAI systems) and everyday uses did not inherently constitute high risk. For example, utilizing ChatGPT for creative endeavors, such as writing poetry or fiction, does not inherently pose a high risk.

Reconciling the inclusion of GPAI into the regulatory framework while maintaining the integrity of the risk-based approach required careful consideration.

A solution then emerged in order to address the challenge posed by the inclusion of powerful GPAI models in the regulatory framework of the AI Act. This solution involved the introduction of an entirely new category called "systemic risk" specifically for potent GPAI models.[10] All other GPAI applications would be placed in the yellow "Low Risk" category, albeit with enhanced obligations. This led to the introduction of an entirely new Title VIII into the Commission's initial draft, set out as follows (numbers of articles are not definitive yet):

The question of whether the term "systemic risk", used for powerful GPAI models, is appropriate in this context is interesting and one that deserves debate. The term "systemic risk" has always been associated, in other contexts, with risks that entail endangering potential outcomes, such as "wide-ranging, cross-sectoral, or transnational impacts". As a United Nations Office for Disaster Risk Reduction publication has highlighted:

"Systemic risk is associated with cascading impacts that spread within and across systems and sectors (e.g. ecosystems, health, infrastructure and the food sector) via the movements of people, goods, capital and information within and across boundaries (e.g. regions, countries and continents). The spread of these impacts can lead to potentially existential consequences and system collapse across a range of time horizons".

[9] Amendments 99, 169 and 394 of the Parliament's negotiating position (P9_TA(2023)0236).

[10] See Kutterer, C. (2024) 'Regulating Foundation Models in the AI Act: From "High" to "Systemic" Risk, AI Regulation Papers 24-01-1, AI-Regulation.com, January 12th.

Recital 60m (provisional number) of the AI Act tries to justify applying the term "systemic risk" to powerful GPAI models in the following way:

"General purpose AI models could pose systemic risks which include, but are not limited to, any actual or reasonably foreseeable negative effects in relation to major accidents, disruptions of critical sectors and serious consequences to public health and safety; any actual or reasonably foreseeable negative effects on democratic processes, public and economic security; the dissemination of illegal, false, or discriminatory content. Systemic risks should be understood to increase with model capabilities and model reach, can arise along the entire lifecycle of the model, and are influenced by conditions of misuse, model reliability, model fairness and model security, the degree of autonomy of the model, its access to tools, novel or combined modalities, release and distribution strategies, the potential to remove guardrails and other factors. In particular, international approaches have so far identified the need to devote attention to risks from potential intentional misuse or unintended issues of control relating to alignment with human intent; chemical, biological, radiological, and nuclear risks, such as the ways in which barriers to entry can be lowered, including for weapons development, design acquisition, or use; offensive cyber capabilities, such as the ways in vulnerability discovery, exploitation, or operational use can be enabled; the effects of interaction and tool use, including for example the capacity to control physical systems and interfere with critical infrastructure; risks from models of making copies of themselves or "self-replicating" or training other models; the ways in which models can give rise to harmful bias and discrimination with risks to individuals, communities or societies; the facilitation of disinformation or harming privacy with threats to democratic values and human rights; risk that a particular event could lead to a chain reaction with considerable negative effects that could affect up to an entire city, an entire domain activity or an entire community.

We will not delve further into this question, as it will be the subject of a separate article by our colleague, Stavros Tsipras. However, we can conclude here that the transposition of the existing concept of "systemic risk" into the AI Act allowed for the preservation of the "risk-based" approach that inspired the Regulation, and thus the Commission's pyramid. Nevertheless, a new urgent question arises: where should one place 'systemic risk', and more broadly, GPAI, within the Commission's pyramid?

2.2. Where Should Systemic Risk Appear in the Pyramid?

Given the heated debates that characterized the regulation of GPAI models by the AI Act until the very last minute, considerable uncertainty remains regarding whether the Commission will choose to integrate the new "systemic" risk tier into its existing pyramid structure. There is a possibility that the Commission may opt against updating its pyramid to avoid further controversies surrounding the criticality of the risk posed by GPAI models and their regulation under the AI Act.

However, we contend that integrating the "systemic" risk tier into the current pyramid structure would harmonize it with the overarching logic of the AI Act and uphold its conceptual coherence. Therefore, we propose to the academic community a revised pyramid model that incorporates the "systemic risk" tier. We believe that this revised model accurately reflects the underlying principles and provisions outlined in the final text of the AI Act.

Within the MIAI AI-Regulation Chair, we have engaged in significant discussions regarding the precise placement of the "systemic" risk tier within the current pyramid structure. The view has been expressed that the "Systemic Risk" tier should be positioned above the "High Risk" tier, distinguished by a more alarming color. This suggestion takes into account the potential cascading impacts of "systemic" risks and their potentially severe societal consequences if they materialize.

While we acknowledge the merit of placing "Systemic Risk" above (or inside) "High Risk" *in other systems*, we contend that this should not be the case within the AI Act pyramid. We base this stance on several factors, including considerations of criticality, breadth of regulation, and mass, which we will elaborate on below. However, while we find our

classification scientifically coherent and satisfactory, we welcome other perspectives on this matter.

Here, we outline the reasons that led us to position the "Systemic Risk" tier below the "High Risk" tier in our pyramid, with the hope that this will stimulate an engaging academic discourse.

(a) – Criticality

As noted in the AI Act, systemic risks are characterized by high complexity, multiple uncertainties, and major ambiguities.[11] The categorization of "systemic risk" in the AI Act hinges on likelihood and conditional use. The Commission highlights that the capabilities of powerful GPAI models 'are not yet well enough understood,' *potentially* posing systemic risks. Therefore, subjecting their providers to additional obligations is considered reasonable. However, it remains uncertain whether such systemic risks will ever materialize. Indeed, one could hope that the due diligence obligations imposed by the AI Act on developers of potent GPAI models could effectively manage the systemic risks associated with these models through rigorous risk assessment and governance.

In contrast, "high-risk" AI applications are consistently classified as such, without conditional considerations. The Act offers a clear definition of "high-risk" applications and a robust methodology for identifying high-risk AI systems within the legal framework. The classification of "high-risk" is determined by the AI system's **intended purpose**, aligning with existing EU product safety legislation.[12] Products covered by sectorial Union legislation are always considered high-risk when subject to third-party conformity assessment under that legislation.[13] Additionally, Annex III of the AI Act lists AI systems that create adverse impacts on people's safety or fundamental rights, which are also automatically classified as high-risk, unless a provider successfully demonstrates otherwise based on the exemptions outlined in Article 6(2a)ff.[14]

These classifications, along with the discussions surrounding the regulation of GPAI models by the AI Act and the desire to "avoid overregulation", suggest that legislators may have considered placing the potential "systemic risks" posed by such models under the tier of "high-risk" applications of AI systems. It is also important to note that, as demonstrated by our pyramid model, these two categories are **not mutually exclusive**. **Therefore, a "systemic risk" system should also be considered "high risk" if it is used for any of the purposes listed in Annex III**. Our expectation is indeed that several systems built on top of GPAI models will ultimately fall into the high-risk category, due to their intended use.

Furthermore, our conclusion that the "Systemic risk" category should be positioned below the "High Risk" one in the criticality pyramid is strengthened by the disparities in obligations for providers of high-risk AI systems compared to developers of powerful GPAI models under the AI Act. The AI Act imposes more stringent requirements on providers of high-risk AI systems,[15] including the mandatory conduct of a conformity assessment before commercialization, a stipulation absent for GPAI models.[16]

[11] Even before the emergence of GPAI, the Council of the EU stressed the need to address 'the opacity, complexity, bias, a certain degree of unpredictability and partially autonomous behaviour of certain AI systems, to ensure their compatibility with fundamental rights and to facilitate the enforcement of legal rules' (Explanatory Memorandum of the AI Act Proposal, Section 1, point 1.1).

[12] 'The classification of an AI system as high-risk is based on the intended purpose of the AI system, in line with existing product safety legislation' (Explanatory Memorandum of the AI Act Proposal, Section 5, point 5.2, para. 5.2.3).

[13] See T. Karathanasis, Guidance on Classification and Conformity Assessments for High-Risk AI Systems under EU AI Act, AI-Regulation.com, February 22th, 2023.

[14] As we will discuss below, an important **exception** has been introduced in the final AI Act to cover cases where a provider considers that an AI system referred to in Annex III is not high-risk.

[15] Before placing a high-risk AI system on the EU market or otherwise putting it into service, providers must subject it to a conformity assessment. This will allow them to demonstrate that their system complies with the mandatory requirements for trustworthy AI (e.g. data quality, documentation and traceability, transparency, human oversight, accuracy, cybersecurity and robustness). For biometric systems a third-party conformity assessment is always required. Providers of high-risk AI systems will also have to implement quality and risk management systems to ensure their compliance with the new requirements and minimise risks for users and affected persons, even after a product is placed on the market.

[16] Providers of models that pose systemic risks are mandated to **assess and mitigate risks**, **report serious incidents**, **conduct state-of-the-art tests and model evaluations**,

(b) – Quantity

As explained earlier, the pyramid wishes to serve a dual purpose beyond illustrating criticality; it also aims to depict the breadth of AI applications subject to new EU regulations.

Placing "Systemic Risk" under the "High Risk" category thus created an anomaly that we tried to correct by reducing the "mass" of the "Systemic Risk" level. The anomaly is the following:

As will be explained later, under the AI Act, a GPAI will be presumed to pose a systemic risk only where the cumulative amount of computing power used to train it was greater than 10^25 floating point operations (flops). As the Commission notes today probably only OpenAI's GPT-4 and likely Google DeepMind's Gemini pass this threshold.

Placing the "Systemic Risk" level under the "High Risk" one, for the reasons of criticality and importance of regulation explained above, while maintaining a "normal" height and mass for this level, could have thus given the impression that the number of "Systemic Risk" systems is superior to the number of "High Risk" applications. Taking into consideration that, for the time being, only two (albeit important) models seem to enter into the "Systemic Risk" category, while several "High Risk" applications appear in Annexes II and III, this would have been misleading.

This is the reason why, in order to address this anomaly, we have substantially reduced the height and mass of the "Systemic Risk" level in our pyramid.

In reality, the exact height and mass of the "Systemic Risk" level in the future will depend on the scientific progress and the threshold. The AI Office (established within the Commission) may update this threshold in light of technological advances, and may furthermore in specific cases designate other models as such based on further criteria (e.g. number of users, or the degree of autonomy of the model). It is interesting to note, for instance, that while the AI Act has not yet even been adopted, countries like France seem to call for a revision of the threshold above the 10^25 flops limit…[17]

2.3. From Menkaure to Khufu?

The Commission's pyramid is thus not only standing, but probably became bigger at the end of the process. Indeed, one could notice that the Commission's pyramid, which was reminiscent of the robust structure of Menkaure, has experienced substantial expansion due to numerous additions proposed primarily by the Parliament and subsequently accepted by the Council during the final compromise. Therefore, from the modest proportions of the Menkaure pyramid, we have transitioned to a larger structure akin to the pyramid of Khafre, if not Khufu.

The principal additions to the initial AI Act include:

First and foremost, the new "systemic risk" tier, that we have explained above.

Secondly, the previously sparse "red" layer has witnessed enlargement due to the introduction of four new AI applications that are consistently deemed to pose an "unacceptable risk," along with tightened restrictions on real-time biometric identification exceptions. Table 1, annexed at the end of this article, permits to identify clearly the new additions.

Thirdly, the radiant "yellow" layer has undergone significant enlargement, encompassing all non-systemic risk GPAI applications with enhanced obligations (article 52c in the current draft), including adherence to Union copyright law.

Other important additions to the initial text proposed by the Commission also took place, especially in the sections concerning enforcement and supervision.

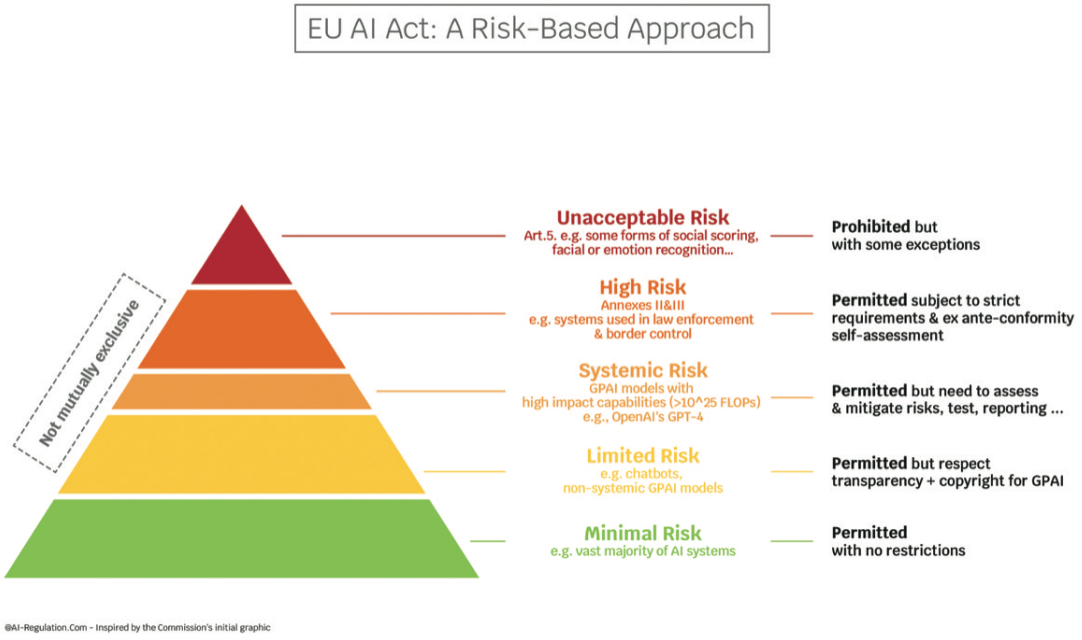
In the last part of the article we will give some more details about the main additions.

ensure **cybersecurity** and provide **information on the energy consumption** of their models.

[17] See for instance here.

3) The Visualization Pyramid: A Closer Examination

We will now present our visualization pyramid in detail.



3.1. Minimal/No Risk (Green Layer)

AI systems that do not fall under one of the four main risk tiers are classified as minimal/no-risk. It is to be expected that a great number of AI applications will fall under this category.

The EU AI Act allows for minimal-risk AI systems to be freely used, while voluntary codes of conduct are encouraged. The codes can be created within organisations or by industry bodies, to be followed by member organisations.

It is crucial to emphasize that existing EU legislation, such as the GDPR concerning the processing of personal data, applies when the use of AI systems falls within the scope of such pre-existing regulation.

3.2. Unacceptable Risk (Red Layer)

This category includes all those AI systems whose use is considered unacceptable as they contravene 'Union values of respect for human dignity, freedom, equality, democracy and the rule of law and Union fundamental rights, including the right to non-discrimination, data protection and privacy and the rights of the child'.[18] Certain forms of system, used for the following purposes, are prohibited:

[18] Recital 15 AI Act Provisional Agreement.

- **Social scoring** for public and private purposes;
- **Exploitation of vulnerable persons via the use of subliminal techniques**;
- **Real-time remote biometric identification in publicly accessible spaces by law enforcement**, subject to a number of exceptions;
- **Biometric categorisation** of natural persons based on biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs or sexual orientation;
- **Individual predictive policing**;
- **Emotion recognition in the workplace and education institutions,** unless for medical or safety reasons (e.g. monitoring how tired an aircraft pilot is);
- **Untargeted scraping** of the internet or CCTV for facial images to build up or expand databases.

The following four series of observations may be made about the changes that came about during the negotiation process:

(a) - Subliminal Techniques and Exploitation of Vulnerable Persons

Initially, the Commission's proposal focused only on AI systems that 'deploy subliminal techniques beyond a person's consciousness'.[19] In the final version of the AI Act, the wording has evolved somewhat, adding 'purposefully manipulative or deceptive techniques' to this prohibited category.[20] The revised causal link between the use of the manipulative technique and the harm, as provided for in the AI Act following the Parliamentary approach, provides two criteria that help assess whether an AI system significantly alters human behaviour.[21] While the intention may be to address well known manipulative techniques such as subliminal messaging, it inadvertently limits the prohibition to cases where both criteria are met and convincingly demonstrated, further complicating the task of proving causation between the use of AI and the resulting harm.[22]

While the prohibition of AI systems that exploit the vulnerability of a specific group of persons[23] may sometimes coincide with the previous one, it targets practices that don't necessarily rely on subliminal methods, but rather exploits the diminished autonomy of certain individuals. In the Commission's proposal, this pertains particularly to children, owing to their age, and to individuals who are mentally or physically disabled. Both the Council and the Parliament have suggested broadening the scope of the prohibition.[24] However, the Council approach prevailed in the finalised AI Act. An illustration of such AI systems can be seen in AI-powered chatbots that engage with children or individuals who are experiencing emotional or psychological distress, or both.

It is important to emphasise that the prohibitions on manipulative and exploitative practices outlined in this Regulation do not impinge upon lawful practices used in medical arenas, such as therapeutic treatment for mental illness or physical rehabilitation. These practices must adhere to the relevant legislation and medical standards, including obtaining explicit consent from individuals or their legal representatives.

[19] Article 5§1 (a) of the AI Act proposal.
[20] The proposal to add such techniques to the list of prohibited AI practices stems from the European Parliament's negotiating position of June 2023. Even though Recital 16 equates subliminal techniques with 'manipulative techniques' that are used 'to persuade persons to engage in unwanted behaviors, or to deceive them by nudging them into decisions in a way that subverts and impairs their autonomy, decision-making and free choices'. The addition by the Parliament complemented a term ('subliminal techniques') that has nevertheless remained undefined in the text.
[21] i) the AI must significantly impair an individual's ability to make informed decisions; ii) as a result, it must lead the individual to make choices he or she wouldn't have made otherwise.
[22] Casaburo, D. & Gugliotta, L. (2023) 'The EU AI Act proposal(s): Manipulative and exploitative AI practices'. Ku Leven CiTiP, blog, September 22nd.
[23] Article 5§1 (b) AI Act.
[24] The Council proposed including a person's 'specific social or economic situation' as a vulnerability factor. This addition has been praised for its focus on systems that exploit individuals' financial struggles or socioeconomic status in order to influence their decisions. On the other hand, the Parliament further proposed 'characteristics of […] individual's or group of persons' known or predicted personality traits', while adopting the Council's amendment.

(b) - Biometric Categorisation or Identification, Face and Emotion Recognition

The ban on real-time biometric identification[25] for law enforcement purposes has been the subject of considerable debate within European institutions, revealing a wide spectrum of opinion on exemptions from prohibitions vis-à-vis AI systems in publicly accessible areas.

The Commission initially proposed[26] a ban on the use of real-time biometric identification for law enforcement purposes, save for three specific exceptions: to locate victims of crime, including missing children; to avert imminent threats such as terrorist attacks; and to identify and locate persons facing criminal charges that carry a minimum sentence of three years' imprisonment.[27]

The EU Council extended the exceptions that relate to law enforcement.[28] Also, according to the Council, publicly accessible spaces should not include prisons or border control areas.[29] Real-time biometric remote identification by officers in these locations would have therefore been allowed.[30]

The text proposed by the European Parliament however took a different approach to that of the Council of the EU, advocating a complete ban on the use of real-time remote biometric identification in publicly accessible spaces.[31] Contrary to the Commission's proposal, the Parliament's negotiating position also envisaged extending the ban to include ex-post use (except in cases involving serious crime and pre-trial authorisation).

Finally, the Commission's proposal prevailed in the finalised AI Act with real-time biometric identification for law enforcement purposes being prohibited, unless these systems are used for any of the listed specific purposes, such as searching for victims of human trafficking or sexual exploitation, or for the prevention of terrorist attacks. In principle, relying on such an exception will require thorough assessments, technical and organisational measures, notifications and a warrant.

Certain AI systems that involve the processing of biometric data also face an outright ban under the finalised AI Act. For example, AI-driven biometric categorisation systems that derive an individual's race, political views, trade union membership, religious beliefs, sexual orientation or other sensitive information from biometric data are prohibited, save for limited exceptions that relate to law enforcement. The European Commission considered biometric categorisation to be a 'high-risk' AI system,[32] while the Council of the EU did not perceive biometric categorisation systems to be high-risk AI systems and only imposed transparency obligations on them. However, the Parliament's position prevailed, considering that they posed an unacceptable risk, and banned them, save for certain exceptions that pertain to therapeutic services.

In addition, the AI Act prohibits the use of AI systems that indiscriminately collect facial images from the internet (web-scraping) or CCTV to create or expand facial recognition databases.[33] AI systems used to infer emotions in workplaces or educational institutions are also prohibited, except for when this is carried out for medical or security reasons.[34]

[25] Article 5§1 (d) AI Act.
[26] For a detailed presentation and a visualisation table see T. Christakis, M. Becuywe & AI-Regulation Team, "Facial Recognition in the Draft European AI Regulation" Final Report on the High-Level Workshop Held on April 26, 2021, AI-Regulation.com, May 27, 2021.
[27] AI Act proposal.
[28] In particular: law enforcement, border control, immigration or asylum authorities would have been allowed to use the relevant systems, in accordance with EU or national law, to identify a person, even against his or her will, 'who either refuses to be identified during an identity check or is unable to state or prove his or her identity' (Recital 19 of the Council's General Approach).
[29] Recital 9 of the Council's General Approach.
[30] It should be stressed that these discussions coincided with the French Parliament's proposal to introduce facial recognition technology in public areas for the Paris 2024 Olympic Games. See Lodie, A. & Celis Juarez, S. (2023) 'AI-Assisted Security at the Paris 2024 Olympic Games: From Facial Recognition to Smart Video', AI-Regulation.com, January 27th.
[31] Parliament's negotiating position (P9_TA(2023)0236).
[32] Article 5§1 (ba) AI Act Provisional Agreement.
[33] Article 5§1 (db) AI Act Provisional Agreement.
[34] Article 5§1 (dc) AI Act Provisional Agreement.

(c) - Crime 'Prediction'

AI-based crime 'prediction' is being increasingly used by European law enforcement and criminal justice authorities to profile people and areas, predict supposed future criminal behavior or occurrence of crime, and assess the alleged 'risk' of offending or criminality in the future. Europol's Innovation Lab is using, for example, AI systems to process massive amounts of data to identify trends and patterns, as well as leveraging tools such as ChatGPT to act as investigative assistants.

The ban on 'predictive' policing and crime prediction systems[35] is significantly weaker than the version voted for by the European Parliament in June in terms of scope, exclusions, clarity and enforcement. Indeed, the finalised text specifically targets the use of AI systems for risk assessment in relation to criminal offences only (narrow scope). This leaves room for the potential use of AI systems for other harmful purposes not directly related to criminal activities, such as discrimination in employment or financial services based on profiling or personality traits. In addition, the second text includes a carve-out that exempts AI systems used to assist human assessment of involvement in criminal activity based on objective and verifiable facts (exclusions). This exception potentially undermines the prohibition by allowing AI systems to be used indirectly to make risk assessments in relation to criminal activity, as long as they support human assessments. In addition, the second text is quite ambiguous in defining what constitutes "objective and verifiable facts directly linked to a criminal activity" (clarity). This ambiguity may lead to different interpretations and potential loopholes in the ban, allowing the continued use of AI systems in ways that could still pose risks to individuals. Finally, by focusing solely on the use of AI systems to assess or predict the risk of committing criminal offences, the prohibition overlooks other potential harms that AI systems could cause, such as invasion of privacy or exacerbation of societal biases and inequalities. As a result, enforcement efforts may be less effective in addressing these broader concerns (limited enforcement).

(d) - Social Scoring

This prohibition extends to what are commonly referred to as 'social scoring' practices.[36] These AI-driven methods determine access to social benefits and/or allocate treatment based on an individual's assessed behavioral patterns or attributes (e.g., socioeconomic status, ethnicity[37]). For instance, a system that identifies vulnerable children in need of social assistance might base its assessment on the trivial or inconsequential behaviors of the parents, such as missing medical appointments or undergoing a divorce.

Initially, the Commission's proposal prohibited social scoring practices that are adopted 'by public authorities or on their behalf'. The phrase 'by public authorities or on their behalf' was deleted in the finalised version of the AI Act, with all social scoring practices being banned, regardless of the user. Such practices are therefore prohibited, since they may lead to discriminatory outcomes or the exclusion of certain groups and therefore, violate the right to dignity and non-discrimination and the values of equality and justice. In order not to prevent the use of AI systems that assist in the legitimate implementation of government (social) policies, these practices are prohibited only if, i) based on data collected in a given domain, they are used to discriminate against people in other, unrelated domains; and/or ii) they simply lead to 'unjustified or disproportionate' treatment of people.

3.3. High-Risk (Deep Orange Layer)

The second tier of regulation pertains to AI applications deemed to pose a substantial risk to people's safety or fundamental rights. High-risk AI systems are permitted under specific conditions, provided they adhere to a set of predefined guidelines. As initially suggested, high-risk AI applications that pose significant risks would be obliged to comply with a comprehensive set of

[35] Article 5§1 (da) AI Act Provisional Agreement.
[36] Article 5§1 (c) AI Act.
[37] Raz A. (2023) 'AI-driven risk scores: should social scoring and polygenic scores based on ethnicity be equally prohibited?', Front. Genet. - ELSI in Science and Genetics, Vol. 14, May 30th.

requirements.[38] The final negotiations sought to yield a greater degree of clarity and feasibility, thereby reducing the burden on stakeholders. Amendments were made to ensure technical feasibility and alleviate burdens.[39]

The AI Act considers two types of AI systems to be high-risk.[40] The first type concerns AI systems that are to be used as a product (or the security component of a product) and are covered by specific EU legislation, such as products used in civil aviation, vehicle security, marine equipment, toys, lifts, pressure equipment and personal protective equipment.[41] The second type concerns AI systems listed in Annex III, such as remote biometric identification systems, AI used as a safety component in critical infrastructure, and AI used in education, employment, credit scoring, law enforcement, migration and the democratic process.[42] Examples of high-risk AI applications encompass the management of critical infrastructure, educational processes such as student allocation and assessment, employment and workforce management, access to public and private services, law enforcement, and migration and border control, among others.

The most notable change from the Commission's initial proposal, as endorsed in the earlier negotiation stages, involves the **creation of an exemption** to this classification. If an AI system that is categorised under the second tier of high-risk AI systems (as outlined in Annex III) does not pose a significant risk of harm, to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making, it will not be classified as a high-risk AI system.[43] In practice, this exception will hold significant weight, as many AI system providers will likely seek to claim that their systems do not pose such risks.[44] This is so that they can circumvent the substantial regulatory obligations and costs associated with high-risk AI classification. However, if a provider intends to invoke this exception, it must thoroughly document its assessment in order to demonstrate compliance.

Regarding the prohibition of AI practices that entail unacceptable risks, exceptions to this prohibition exist, as previously mentioned, for specific purposes outlined in the legislation, such as searching for victims of human trafficking or sexual exploitation, or the prevention of terrorist attacks. Whenever one of these exceptions applies, the AI systems fall automatically under the high-risk tier. Indeed, when their use is not banned under Article 5, AI systems used for biometric identification, categorisation, or emotion recognition should be classified as 'high-risk' and be subject to stringent regulations, including extensive assessments, the implementation of technical and organisational measures, notification procedures and the obtaining of a warrant. Biometric systems solely utilised for cybersecurity and personal data protection fall outside the scope of regulations that apply to 'high-risk' AI systems.

3.4. Systemic Risk (Orange Layer)

The regulation of so-called General Purpose AI (GPAI)[45] models, such as GPT-4, which was first introduced in the EU Parliament's negotiating position in June 2023,[46] was fiercely debated in the final stages of the trilogues and deemed particularly

[38] These encompass the establishment of a comprehensive risk management framework, adherence to data usage and governance protocols, maintenance of detailed record-keeping mechanisms as outlined in the finalised AI Act, provision of transparent information to users, implementation of human oversight mechanisms, and ensuring adequate levels of cybersecurity, robustness, and accuracy (Title III of the AI Act).

[39] Such as modifications to obligations regarding the quality of training data and adjustments to technical documentation requirements for small to medium-sized enterprises (SMEs).

[40] See Karathanasis, T. (2023) 'Guidance on Classification and Conformity Assessments for High-Risk AI Systems under EU AI Act', AI-Regulation.com, February 22nd.

[41] Article 6§1 AI Act.

[42] Article 6§2 AI Act.

[43] This exemption applies if the system is designed for (a) narrow procedural tasks, (b) enhancing the outcomes of previously completed human activities, (c) identifying decision-making patterns or deviations from past decisions without directly replacing or influencing human assessments without appropriate human review, or (d) undertaking preparatory tasks for an assessment (Article 6§2a AI Act).

[44] However, if the AI system engages in the profiling of individuals, it is consistently considered a high-risk AI system.

[45] A GPAI is an 'AI model, including when trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable to competently perform 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' (Article 3 (44b) AI Act).

[46] European Parliament's negotiating position (P9_TA(2023)0236).

controversial due to fears that excessive regulation would hinder innovation and harm European companies. It should be remembered here that the *GPDP v. Open AI* case of March 2023[47] had been the source of a significant domino effect[48] among European Data Protection Authorities regarding the regulation of similar AI models,[49] in relation with potential violations of several articles of the GDPR. Following prolonged debates aimed at breaking the deadlock on foundation models, the EU Parliament and Council reached a compromise in the form of an amended tiered approach, which involved a shift away from 'very capable' foundation models to 'high impact' GPAI models, to highlight the focus on the systemic risks these models can pose.

The AI Act now identifies two categories of GPAI models: generic GPAI models and "systemic" GPAI models. The introduction of general transparency obligations *for all GPAIs* and specific obligations for systemic GPAIs underlines a broader regulatory scope, incorporating measures to mitigate foreseeable risks and enhance transparency. This approach aligns with broader legislative efforts to manage the impacts of digital technologies, as seen in the Digital Services Act (DSA) and Digital Markets Act (DMA), indicating a nuanced understanding of AI's potential and the importance of proactive risk management without fundamentally departing from the risk-based regulatory framework.

The term "systemic risk"[50] first appeared in the compromise text of December 2023, since neither the Commission nor the European Parliament/Council had previously made any reference to such a term in their respective documents. It has been acknowledged, for example, that 'powerful models could cause serious accidents or be misused for far-reaching cyberattacks. Many individuals could be affected if a model propagates harmful biases across many applications'.[51]

According to Article 52a§2 of the finalised AI Act, a GPAI will be presumed to pose a systemic risk where the cumulative amount of computing power used to train it was greater than 10^25 floating point operations (flops).[52] Over time, the threshold may be revised by the AI Office based on considerations such as the number of parameters in the model, the size of its training data set, characteristics relating to how it functions and its number of users, to accommodate technological advancements and shifts in industry standards, including enhancements to algorithms or advances in hardware efficiency.[53]

Regarding the obligations on providers of systemic GPAI models, such models are subject to the same obligations as basic GPAI models,[54] along with **additional** requirements that collectively contribute to more extensive regulation.[55] The AI Act provides that, pending the publication of harmonized European standards, GPAI models that pose a systemic risk may rely on "codes of practice"[56] to demonstrate compliance with their obligations.[57]

[47] Garante per la Protezione dei Dati Personali (GPDP), Provvedimento n. 112 del 30 marzo 2023 [9870832].
[48] BEUC (2023), 'Investigation by EU authorities needed into ChatGPT technology', Press release 30 March 2023.
[49] The Italian ChatGPT saga provided the first relevant case in the EU concerning raising GDPR issues in relation to the deployment of LLMs. Among the reported violations, the following were mentioned: lawfulness, transparency in the use of data subjects' personal data, rights of the data subjects, children's personal data processing, and data protection by design and by default. For details, watch the video that our panel organised during the IAPP AI Governance Global Conference, Boston, on November 2, 2023: Can Generative AI Survive the GDPR?.
[50] A risk that is specific to the high-impact capabilities of general-purpose AI models, that has a significant impact on the internal market due to its reach, and has actual or reasonably foreseeable negative effects on public health, safety, public security, fundamental rights, or society as a whole, that can be propagated at scale across the value chain (Article 3 (44d) AI Act Compromised Text).
[51] European Commission, Artificial Intelligence – Questions and Answers, Brussels, 12 December 2023.
[52] Cumulative amount of computation used to train the GPAI model (Recital 60n AI Act Provisional Agreement).
[53] Annex IXc and Recital 60n AI Act Provisional Agreement.
[54] Article 52c§1 AI Act Provisional Agreement.
[55] Specifically, they are required to: a) conduct model evaluations using standardized protocols and tools that reflect current advancements, including the execution and documentation of "adversarial tests" aimed at identifying and mitigating systemic risks; b) evaluate and mitigate potential systemic risks at the Union level, including their origins, stemming from the development, market introduction, or utilization of generalized AI models that carry systemic risks; c) monitor, document, and promptly report the relevant information on significant incidents and potential corrective actions to the AI Office and, where applicable, to competent national authorities and; d) ensure adequate cybersecurity measures pertaining to the model and its physical infrastructure (Article 52d and Annex IXa (Sect. 2) AI Act Provisional Agreement).
[56] By "codes of practice" we mean technical documents that report the standards of a technological sector.
[57] Article 52e AI Act Provisional Agreement.

The question at hand is whether this level of regulation could hinder the development of European Foundation Models, a concern voiced by certain governments, notably France. Naturally, the Commission's assessment of these obligations will be crucial. For instance, with regard to "adversarial testing," the discretion of the AI Office in determining the adequacy of the testing process will be significant. Regarding incidents, it will be important to ascertain whether the GPAI model should be capable of reporting all information pertaining to the AI systems that utilise it (given that they are typically managed by different entities).

3.5. Limited Risk (Yellow Layer)

The AI Act imposes transparency obligations on a series of AI systems which do not fall under the previous categories, including AI systems that are intended to directly interact with natural persons (e.g. AI companions); AI systems that generate deep fakes; or AI systems, including non-systemic GPAI systems, that generate synthetic audio, image, video or text content (e.g. Midjourney, DALL-E).

In such cases, users must be notified of the existence of an AI system and be aware that they are interacting with a machine. In certain scenarios, the content must be tagged in a machine-readable format so that it can be identified as being artificially generated or manipulated.[58] The AI Act provides for exceptions to this requirement in certain circumstances, such as when it relates to law enforcement or when the AI system is used for artistic, satirical, creative or similar purposes.[59]

More specifically, these categories will be subject to transparency obligations,[60] which consist of a) guaranteeing that technical documentation that aids understanding of how they work (including documentation concerning the data training process) is available to the AI Office and the national competent authorities,[61] as well as to those third parties who intend to integrate the model into their AI systems (downstream providers);[62] b) having a policy concerning the respect of EU copyright rules,[63] in particular to ensure that, where copyright holders have opted out of allowing their data to be available for text and data mining (including web-scraping), this is identified and respected; c) preparing and publishing a statement about the data used to train the general purpose AI model.[64] Open-source general purpose AI are exempt from the requirements concerning documentation and downstream information but they must have a copyright policy and information must be provided about training data.[65] Regarding the code of practice, non-systemic GPAI may also rely on "codes of practice" to demonstrate compliance with their obligations.[66]

[58] Recital 70b AI Act Provisional Agreement.
[59] Recital 70b AI Act Provisional Agreement.
[60] Article 52c AI Act Provisional Agreement.
[61] Article 52c§1 (a) and Annex IXa (Sect. 1) AI Act Provisional Agreement.
[62] Article 52c§1 (b) and Annex IXb AI Act Provisional Agreement. Notably, the AI ACT addresses instances of vertical integration, wherein the GPAI provider aligns with the deployer of the respective AI system. Here, the AI Office functions as a market oversight body, supplanting the authority typically held by national entities. This marks the initial and conspicuous application of specialised competition regulations within the AI sector, an area where many have voiced concerns regarding consolidation risks. It is noteworthy that the European Commission has initiated an inquiry into the relationship between OpenAI and Microsoft.
[63] Article 52c§1 (c) AI Act Provisional Agreement.
[64] Article 52c§1 (d) AI Act Provisional Agreement.
[65] Article 52c§ -2 AI Act Provisional Agreement.
[66] Article 52c§3 AI Act Provisional Agreement.

Annex (Table 1) Article 5 –Prohibited AI Systems that Pose an "Unacceptable Risk" (Additions)	
(a) the placing on the market, putting into service or use of an AI system that deploys subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective to or the effect of materially distorting a person's or a group of persons' behaviour by appreciably impairing the person's ability to make an informed decision, thereby causing the person to take a decision that that person would not have otherwise taken in a manner that causes or is likely to cause that person, another person or group of persons significant harm	
(b) the placing on the market, putting into service or use of an AI system that exploits any of the vulnerabilities of a person or a specific group of persons due to their age, disability or a specific social or economic situation, with the objective to or the effect of materially distorting the behaviour of that person or a person pertaining to that group in a manner that causes or is reasonably likely to cause that person or another person significant harm	
ADD ->	(ba) the placing on the market or putting into service for this specific purpose, or use of biometric categorisation systems that categorise individually natural persons based on their biometric data to deduce or infer their race, political opinions, trade union membership, religious or philosophical beliefs, sex life or sexual orientation. This prohibition does not cover any labelling or filtering of lawfully acquired biometric datasets, such as images, based on biometric data or categorizing of biometric data in the area of law enforcement
(c) the placing on the market, putting into service or use of AI systems for the evaluation or classification of natural persons or groups thereof over a certain period of time based on their social behaviour or known, inferred or predicted personal or personality characteristics, with the social score leading to either or both of the following: (i) detrimental or unfavourable treatment of certain natural persons or whole groups thereof in social contexts that are unrelated to the contexts in which the data was originally generated or collected; (ii) detrimental or unfavourable treatment of certain natural persons or groups thereof that is unjustified or disproportionate to their social behaviour or its gravity;	
(d) the use of 'real-time' remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement unless and in as far as such use is strictly necessary for one of the following objectives: (i) the targeted search for specific victims of abduction, trafficking in human beings and sexual exploitation of human beings as well as search for missing persons; (ii) the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons or a genuine and present or genuine and foreseeable threat of a terrorist attack; (iii) the localisation or identification of a person suspected of having committed a criminal offence, for the purposes of conducting a criminal investigation, prosecution or executing a criminal penalty for offences, referred to in Annex IIa and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least four years. This paragraph is without prejudice to the provisions in Article 9 of the GDPR for the processing of biometric data for purposes other than law enforcement.	
ADD ->	(da) the placing on the market, putting into service for this specific purpose, or use of an AI system for making risk assessments of natural persons in order to assess or predict the risk of a natural person to commit a criminal offence, based solely on the profiling of a natural person or on assessing their personality traits and characteristics. This prohibition shall not apply to AI systems used to support the human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity;
ADD ->	(db) the placing on the market, putting into service for this specific purpose, or use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or CCTV footage;
ADD ->	(dc) the placing on the market, putting into service for this specific purpose, or use of AI systems to infer emotions of a natural person in the areas of workplace and education institutions except in cases where the use of the AI system is intended to be put in place or into the market for medical or safety reasons.

Theodore Christakis is the Director of the Chair AI Regulation. Professor of International & European Law, University Grenoble Alpes. Member: Institut Universitaire de France; French National Digital Council (CNNum); French National Committee for Digital Ethics (CNPEN); EIT-Health Consultative group on AI, Data, Industrial & Cybersecurity Policy. Senior Fellow with the Cross-Border Data Forum. Director, Center for International Security & European Studies (CESICE); co-Director, Grenoble Alpes Data Institute.

Dr. Theodoros Karathanasis holds a PhD in European law from the Faculty of Law of the University of Grenoble Alpes. He is attached to the Centre for International Security and European Cooperation Studies (CESICE) and his research work in the field of cybersecurity has been funded by the Grenoble Alpes Institute of Cybersecurity. He is a member of the cyber experts network of the European Centre of Excellence for Combating Hybrid Threats (HybridCoE), as well as the EU CyberNet expert network. He is a Research Fellow at the AI-Regulation.com Chair.

Cover Photo : Created by **DALL-E,** prompt by **Theodore Christakis : "** *Create a painting depicting the evolution from the small pyramid of Menkaure to the larger pyramid of Khufu, symbolizing the transition from the initial stages to the advanced development, in the context of the EU AI Act. Incorporate elements that represent the progression of technology and regulation, capturing the transformation from a smaller, simpler structure to a larger, more complex one***"**

To cite this article: T. Christakis, T. Karathanasis, Tools for Navigating the EU AI Act: (2) Visualisation Pyramid, AI Regulation Papers 24-03-5, AI-Regulation.com, March 8th, 2024.

This work has been partially supported by MIAI @ Grenoble Alpes, (ANR-19-P3IA-0003) and by the Interdisciplinary Project on Privacy (IPoP) of the Cybersecurity PEPR (ANR 22-PECY-0002 IPOP).