moz://a

Artificial Intelligence Act
How the EU can take on the challenge posed by general-purpose AI systems

Key points at a glance

- The **question of general-purpose AI (GPAI) systems needs to be addressed in the AI Act since the original Commission draft fails to account for their special nature**; GPAI is already at the core of some of the AI industry's most successful services, provided by some of the industry's most powerful players.
- Within the framework proposed by the Commission, the AI Act's requirements would be imposed on *either* original providers *or* those adapting GPAI to high-risk uses — but **to effectively prevent harm and avoid overburdening individual actors along the AI supply chain, this responsibility should be shared.**
- Imposing the full burden of compliance on those publishing open source GPAI — and therefore creating incentives to shield GPAI models from public scrutiny — could **stifle important safety and security research as well as downstream innovation.**

Overview

As the EU institutions' work on the AI Act progresses, the role of so-called "general-purpose AI systems" (GPAI) is becoming an increasingly contentious topic. While not part of the Commission's original proposal, questions of such systems' definition, their treatment under the AI Act, and whether to include them explicitly in the AI Act in the first place remain unresolved.

Following most proposed definitions of GPAI (e.g., in European Council compromise drafts and the JURI committee's opinion), the term relates to AI systems performing "generally applicable functions" in "a plurality of contexts."[1] This refers to AI systems that are provided without a specific intended purpose; instead they can serve a large number of purposes, including purposes not foreseen or declared by their original providers. Importantly, this concerns a large part of the AI industry.

Current discussions of GPAI tend to focus on large language or large vision models at the frontier of AI research, for example OpenAI's GPT-3 and DALL·E 2 or Google's PaLM. But the proposed definitions would also cover many pre-trained, multi-purpose AI models (for example for object detection) offered as cloud AI services — such as those provided via AWS, Google Cloud, or Microsoft Azure — that already are in widespread use. These, too, often come without a specified intended purpose within the meaning of the AI Act. The Commission's and other proposals risk letting the burden of compliance fall onto SMEs and other actors adapting GPAI systems for downstream use while (unintentionally) relieving some of the world's biggest companies of responsibility for potential harms caused by technology they develop.

The stakes of appropriately regulating GPAI are significant. GPAI systems can have a plethora of different uses — both intended and unintended. They are used widely already and at an increasing rate. Without specifically tackling GPAI, the AI Act cannot be considered 'future-proof', nor will it be sufficiently able to protect people from harm. However, GPAI's special nature presents regulatory challenges that have not yet been fully solved.

Key regulatory challenges

Distributing Responsibility Appropriately

A binary model of imposing key obligations on *either* providers *or* those adapting AI for high-risk use does not work for GPAI. The Commission's original proposal would effectively lead to the former: In line with Art. 28(1), once an entity adapts a GPAI to a specific purpose or *fine-tunes* a model (i.e., training it with additional, context-specific data), it would have to shoulder all responsibilities of the provider. This would

[1] The latest Council compromise draft (October 2022) defines GPAI as an AI system that "that - irrespective of how the modality in which it is placed on the market or put into service, including as open source software - is intended by the provider to perform generally applicable functions such as image and speech recognition, audio and video generation, pattern detection, question answering, translation and others; a general purpose AI system may be used in a plurality of contexts and be integrated in a plurality of other AI systems[.]"; meanwhile, researchers from the Future of Life Institute and University College London propose to define GPAI as an AI system "that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which it was not intentionally and specifically trained."

disproportionately affect SMEs and less tech-savvy enterprises, to the benefit of larger and more powerful actors in the AI industry. The Council's approach would, in effect, lead to a similar outcome, since it gives the original providers of GPAI the opportunity to shield themselves from responsibility by simply excluding potential high-risk uses in their instructions of use or accompanying information. Such incentives to deflect responsibility run counter to the need to address certain risks upstream during the design and initial development stages. Additionally, the latest Council proposal would delegate spelling out critical details of how to address GPAI to the Commission by implementing act. Conversely, other proposals would place the responsibility for compliance entirely on the original providers while placing no additional obligations on those adapting GPAI — with potentially significant consequences.

In most cases, neither original providers nor those adapting or using GPAI for high-risk purposes downstream are well-placed to comply with *all* obligations in the AI Act. Original providers, on the one hand, will not be able to conduct adequate testing without knowledge of downstream uses and the context of use, or to provide documentation on data used for fine-tuning. On the other hand, those adapting or using GPAI may not be able to provide certain technical documentation without access to the trained model, design specifications, or datasets used for original (pre-)training of a GPAI model. It is therefore necessary to divide responsibilities between these actors and assign obligations to those actors best placed to meet them — both in order for the AI Act to live up to the standard of protection it set out to meet and for actors in the supply chain not to be confronted with obligations they can't effectively comply with.

Accounting for open source GPAI

The proposals put forward by the Czech Council Presidency (including the latest draft) and the JURI committee opinion explicitly include open source models in the definition of GPAI — that is, GPAI models (and, in the case of the JURI proposal, even datasets) made freely available to the public under an open license, not as a commercial service. This would mean that those publishing such models, including community-driven projects, would become responsible for complying with all provider obligations. This would disincentivize publishing open source AI models, as open source communities or researchers working on such models would be wary of liability. Instead, it could lead to important AI research and development being further removed from public scrutiny and more concentrated in the hands of large commercial actors.

There is tremendous value in open source research and development. Making GPAI systems available under an open license advances innovation by providing building blocks for free use across the AI ecosystem, including to SMEs and lower-resource

organizations. It also enables critical public-interest research on the safety and trustworthiness of such large and powerful AI systems. Open source communities should not be held solely responsible for complying with the requirements of the AI Act when AI models and datasets are released open source and not as a commercial service.

If released under an open license, much of the information necessary for anyone to comply with the AI Act would be public information, not held privately by the original provider. Open source GPAI should therefore be treated differently from other, proprietary GPAI systems made available as commercial services, with those adapting them for high-risk use becoming providers as foreseen by the framework proposed by the Commission.

A way forward

Addressing downstream risks is critical to ensure that people are protected from harm caused by systems which were adapted to high-risk uses. Where an actor decides to adapt a GPAI system for a specific high-risk purpose, they should shoulder some of the responsibility. It is therefore important that certain requirements are placed on those adapting GPAI, not original providers, where they are best placed to protect the health, safety, and fundamental rights of individuals. For example, this includes testing the AI system under the conditions in which it is meant to be used or providing for adequate human oversight measures that account for the context of use.

Those adapting GPAI will not be able to meet all the obligations imposed on providers, who are generally more technically sophisticated. Many requirements related to the design and development stage of AI systems would be difficult or impossible for them to address. For instance, if an organization uses a pre-trained GPAI model as cloud-based software-as-a-service, the organization will have little to no insight into the data used to (pre-)train it.

We propose a division of responsibilities that takes into account the special nature of GPAI, effectively shields individuals and communities from harm, and does so without overburdening either original providers or those adapting GPAI for high-risk use downstream. The original providers of a GPAI system should meet certain requirements and obligations, like those related to risk management or the technical design and development of the GPAI system, to the fullest extent possible and undergo conformity assessments. Meanwhile issues that emerge from fine-tuning and adapting GPAI systems for high-risk uses should be addressed further downstream. To facilitate downstream compliance , providers should provide instructions to

downstream actors on when they are subject to additional obligations and how these can be met.

Some of the more intricate details of what obligations should be assigned to whom and under what circumstances may be spelled out via implementing act by the Commission after the AI Act is passed into law, as proposed by the latest Council compromise draft. However, should this be the case, it is important that the Commission takes up this task with a clear mandate and operating within a clearly outlined general approach that effectively divides up responsibilities and ensures that all requirements and obligations foreseen by the AI Act are sufficiently met in the case of GPAI systems, too. This includes meeting data governance requirements at each step where new data is ingested into a GPAI system (e.g., during pre-training or fine-tuning), that risks are managed throughout an AI system's development process and lifecycle and if necessary by multiple actors, and that rigorous testing occurs where a specific context of use is determined. Delegating this task to the Commission should *not* be seen as an easy way out or as an opportunity to water down obligations, removed from public scrutiny and under the influence of those organizations with the most resources to lobby the Commission.

Finally, the AI Act should not actively discourage the release of open source GPAI. Instead it should take a proportionate approach that considers both the special nature of the open source ecosystem as well as the fact that open source GPAI is released with more information than its proprietary equivalent, along with, for example, better means to validate provided information and test the capabilities of GPAI models. GPAI released open source and not as a commercial service should therefore be excluded from the scope of the approach outlined above *if* the information necessary for compliance is made available to downstream actors. This could contribute to fostering a vibrant open source AI ecosystem, more downstream innovation, and important safety and security research on GPAI.

Hypothetical Case Study: GPAI in recruitment

As part of its recruitment process, company A asks applicants to provide work-related writing samples. To speed up the review process, the company wants to use a GPAI model designed to process and generate text to generate brief summaries of the writing samples. The GPAI model is provided as a commercial service by an AI company B through a so-called 'application programming interface' (API).[2] To improve

[2] That is, users of the service can deploy the model for specific tasks but have no direct access to the model itself, its training data, or certain technical parameters.

the quality of the summaries, company A 'fine-tunes' the (already pre-trained) model using writing samples and summaries from previous applicants.

Under the AI Act, this would be considered a high-risk use of AI. If company A were considered the provider due to adapting the GPAI model for high-risk use, it would need to comply with all requirements and obligations under Title III, Chapters II and III. However, it would be difficult for company A to fully comply with, for example, Article 10 (data and data governance) or Article 11 (technical documentation) as it is likely to have no access to testing, training, and validation datasets used by company B to pre-train the model, or to information about design specifications or the model's technical architecture. In fact, both may contain trade secrets of company B.

If company B were solely responsible for complying with the obligations imposed on providers, it could also lack certain information that is critical for compliance. For instance, company B is unlikely to know how exactly company A intends to use their GPAI model. Without this contextual knowledge, and given the myriad potential uses of the GPAI model, it would be difficult to carry out adequate testing of accuracy and robustness or to devise adequate human oversight measures for this specific use context. Further, company B would not have access to the dataset used by company A for fine-tuning the GPAI model, making it impossible at the time of undergoing the conformity assessment to fully comply with Article 10.

Company A and company B should therefore share the responsibility of ensuring effective compliance with the AI Act — and of protecting individuals applying for positions at company A from unfair treatment. Company B, the original provider, should ensure compliance with all requirements and obligations it *can* meet to the fullest extent possible and undergo the conformity assessment. Where company B cannot effectively comply (due to a lack of information about, for example, the exact intended use of the system), it should provide company A, the actor adapting the GPAI system for high-risk use in a recruitment context, with the information necessary for compliance. Company A, in turn, should, amongst other things, ensure adequate human oversight of the system throughout the recruitment process and conduct testing of the system under (near-)real conditions.

For more information, please contact Maximilian Gahntz (max@mozillafoundation.org) and Claire Pershan (claire@mozillafoundation.org).