moz://a

The AI Act and Foundation Models: Considerations for Trilogue

August 10, 2023

Dear MEP Benifei,
Dear MEP Tudorache,
Dear Shadow Rapporteurs,

In a letter from April of this year, you have rightly pointed to the importance of dedicated rules for foundation models and other general-purpose AI systems in the AI Act. Specifically, you committed to developing "a set of rules specifically tailored to foundation models." If developed and deployed responsibly, foundation models and generative AI have significant potential to be used across sectors, be it to spur human creativity or to automate tedious tasks. However, as we and others noted earlier this year in a brief with five considerations to guide the regulation of foundation models in the AI Act — co-signed by more than 60 international AI experts and 12 institutions — such AI models also carry inherent risks to people's rights and safety. These need to be met with robust safeguards. We thus welcome your resolve in this matter and are pleased that you delivered on this commitment in the compromise adopted by the European Parliament

But the work isn't done yet. To make foundation models made available on the EU single market more trustworthy, finding a compromise in trilogue that balances innovation with protecting people from harm is of utmost importance. A tailored and robust framework to curtail the risks posed by foundation models is necessary. In trilogue, it thus is important to protect a number of hard-won accomplishments and not to dilute the safeguards included in the European Parliament's compromise. Specifically, the following points should not be up for negotiation:

- **Upstream obligations for developers and due diligence along the value chain:** While some risks emerging from the use of foundation models are dependent on the context of use and need to be addressed at the application layer, other risks can best — or only — be addressed upstream by these models' original developers. Due diligence and information sharing are thus necessary across the entire value chain, starting at the top.

- **No loophole for original developers:** Original developers of foundation models and other general-purpose systems should not be able to relinquish responsibility for the technology they develop by simply precluding high-risk use of the technology with a boilerplate legal disclaimer, as proposed in the Council's general approach.
- **An expansive definition:** The AI Act should not focus too narrowly on the large language and computer vision models that have recently captured the public's imagination, but also cover any general-purpose AI system that does not come with one specific intended purpose.

Further, the trilogue also offers an opportunity to refine some of the details included in the European Parliament's position and to further strengthen the framework Parliament developed to address the challenges posed by foundation models within the context of the AI Act. Specifically, there are several points worth considering in this context:

- **Open source AI:** The AI Act should account for the special nature and benefits of AI development under free and open source licenses (as well as other permissive licenses). As we at Mozilla and, respectively, a coalition of organizations from the open source community have recently argued, the AI Act should therefore impose proportionate obligations on the developers of open source AI systems, tailored to their capabilities and the specific context of releasing AI under permissive licenses.[1] Otherwise, the AI Act could risk privileging proprietary AI and discourage open source AI research, development, and innovation.
- **Definition:** Defining foundation models as models "trained on broad data at scale" might focus too narrowly on data as one key variable and neglect the role of model size (i.e., the number of parameters in a model) and, as a function of these two, the compute used to train such a model. In fact, there are different configurations of such models designed "for generality of output": from smaller models trained on large amounts of data to models that are bigger in size but trained on less data. It would also leave the difficult task of determining a threshold for what qualifies as "broad data at scale" and what does not.
- **Benchmarks:** In imposing obligations on providers of foundation models, the European Parliament proposal critically relies on benchmarks as a tool for accountability. However, the state of the art in this area is fluid and benchmarks can become rapidly outdated. Further, existing benchmarks are far from covering

[1] It is important to note here that we do not mean this to exempt Mozilla from obligations imposed by the AI Act in the context of our own commercial activities. Rather, we believe that there are other organizations in the open source community, such as small-scale and community-driven open source AI projects, that should be shielded from being overburdened by compliance obligations.

the breadth of risks associated with foundation models and are often developed by such models' developers themselves. To ensure that benchmarking can serve as a robust accountability mechanism and that guidance holds up to scrutiny, the AI Office needs to be sufficiently resourced and should ensure that independent benchmarking and evaluation experts — i.e., experts without current industry affiliation or vested interests — are consulted in the process of developing guidance.

- **Independent experts:** It is a welcome step that the European Parliament compromise prescribes the involvement of independent experts in risk identification and mitigation as well as in evaluation and testing of foundation models. However, without specifying criteria for independence or the nature of experts' involvement, this risks turning this addition into a performative check-box exercise.

Taking these considerations into account during trilogue can help you deliver on your promise: that is, that the AI Act includes effective and future-proof guardrails for foundation models — so that the AI Act enables future innovation in this area of research and development that comes to the benefit, not at the cost, of people. As a public interest-driven and trusted voice in the tech sector, Mozilla stands ready to support these efforts.

Sincerely,

Maximilian Gahntz

Senior Policy Researcher
Mozilla