Brussels, 24 October 2023

Statement on:

Regulation of Foundation Models / General Purpose AI in the EU AI Act

Combined statement of the **European AI Forum EAIF** together with **Aleph Alpha**, **Lengoo** and **Nyonic** on the proposal of the Spanish Council Presidency to regulate foundation models and general purpose AI in the EU AI Act.

Background

The Spanish Presidency of the Council of the European Union has circulated a report ("**Report**") in preparation of the fourth trilogue negotiation on the EU AI Act, taking place on 24 October 2023. The Report proposes a three-tiered regulatory approach for Foundation Models ("**FM**") and General Purpose AI ("**GPAI**"). Tier 1 shall include certain transparency requirements for all FMs. Tier 2 with additional requirements shall apply to so-called *'very capable foundation models'*, to be defined based on computing power measured in FLOPS (floating point operations per second). Tier 3 shall cover certain wide-spread GPAI systems to be defined by thresholds measuring the number of registered business or retail users.

While we is not opposed to a multi-tiered approach to regulate FMs and GPAI, we believe the proposal is not sufficiently in line with the risk-based and use case-specific concept of the AI Act and due to referencing of specific metrics will be neither practicable nor flexible enough based on the following reasons:

Assessment

As the Report rightly points out, there is currently no widely accepted definition of very capable FMs. Differentiating between Tier 1 FMs and Tier 2 FMs (very capable FMs) based solely on computing power (FLOPS) is not sufficient for identifying critical FMs that would require stricter regulation in accordance with the risk-based approach of the AI

Act. This is mainly due to the fact that the number of FLOPS required to train the final version of the model (generally related to CPU usage, not GPU usage) has no impact on the potential risk it poses. There is no correlation between a certain number of FLOPS and the "capabilities" of FMs, other than that it may be "large" (as in a large language model).

Further, the reference to a specific metric leads to the definition in need of constant updating due to technological developments and allows for easy circumvention of responsibilities. **Consequently, we reject the proposed definition of very capable FMs, since neither adequate, nor practical or future-proof.**

The envisaged requirements for all FMs in Tier 1, namely documentation of the model, the training process, the results of internal red-teaming as well as performing and documenting model evaluation based on benchmarks, are of a general nature and not further defined. Consequently, it is not clear from the report how these requirements shall be implemented in practice. Furthermore, aspects of intellectual property and business secrets have not been addressed yet in the transition from pre-market to post-release responsibilities. Further, acknowledging that very capable FMs can pose increased risks, the intended external red-teaming process by vetted red-testers, similar to the concept of so-called trusted flaggers, is not sufficiently clear. **Consequently, we call for a clear definition of requirements and transparent standards for practical implementation to create a level playing field for all market participants.**

In general, we are concerned about the additional bureaucratic burden that this approach would impose on AI companies, as various additional requirements would significantly increase operating costs and slow down the timeline for AI companies to enter the market. **We therefore urge EU lawmakers to strike a balance between effective regulatory requirements and the ability to maintain a competitive and innovative environment for AI companies.**

Proposal and Recommendations

Increasingly powerful foundation models, commercially and open source, are not released in a transparent and open manner (open AI), but as a black closed approach (closed AI). At the same-time we are experiencing ‘open-washing’ and an increasing tension between power concentration upon a few FMs and mitigation of (systemic) risk. How to evaluate FMs and mitigate risks in a targeted manner is currently an unsolved

question, which can be mainly attributed to the lack of transparency in relation to the code as well as to the risk analysis components (e.g. decision making process, data, training process, fine-tuning, evaluations, etc.).

In order to enable an understanding of how FMs should be assessed and how risks can be mitigated, in addition to promoting transparency based on an open and transparent approach and controlling the risks of systems that follow a closed black box approach, and to enable responsible innovation along the AI value chain, we propose the following approach to the European Commission, the Members of the European Parliament and the Council of the European Union for consideration in the trilogue negotiations on the EU AI law and as a possible middle ground for the regulation of FMs:

For all FMs:

- Documentation requirements:
    - **Documentation of the model**
    - **Documentation of the risk analysis components**, i.e. (i) parts of the model development that can provide further insight into the model and its capabilities; decision making process, on what data was collected and how, and documentation of the process, (ii) details on model risks, (iii) training data, fine-tuning data, and information on humans involved in adapting the model through methods such as reinforcement learning with human feedback as well as (iv) evaluation results of any evaluations that researcher and developers may have run on the model in accordance with standardised protocols and tools (i.e. benchmarks), (v) documentation of environmental impact
    - **Documentation of the replication components**, i.e. meaning a technical paper detailing the model training process and code used to train the model, training information such as configuration settings (e.g. batch size), and telemetry collected during training (e.g. training loss).
- **Internal red teaming**, whereas we suggest an approach to be developed by the competent authority (AI Office) with stakeholders. The EAIF offers its collaboration and contribution to support in such development.
- **Collaboration with authority** and upon alert disclosure of information subject to safeguards for intellectual property and business secrets for inspections.

FMs, that are transparent and publish, subject to issues of intellectual property rights and maintaining business secrets (open FMs),

- the model
- the documentation on the risk analysis components
- a responsible use guide for down-stream providers

no further obligations shall apply.

FMs, that do not comply with the transparency requirements above (closed FMs):

- Before market entry:
    - regular external red-teaming through vetted red-testers (to be vetted by the AI Office), with a view to uncover vulnerabilities and identify areas for risk mitigation, the results of which need to be submitted to the Office + introducing a risk assessment and mitigation system, also covering possible systemic risks.
- After market entry:
    - regular compliance controls organised by the AI Office and carried out through independent auditors.
- Documentation sharing:
    - Obligation to provide information on the model and the risk analysis components as well as responsible use to downstream providers.

For more information:

European AI Forum: eaiforum.org
Press and Media: info@eaiforum.org

About the European AI Forum:

The European AI Forum (EAIF) is the first not-for-profit organisation representing the European AI ecosystem. We are designed by and for Europe's AI community and aim to serve as a resource and forum for education, information sharing and networking between companies, policymakers and the general public.