27 November 2024

COMMENTS IN RESPONSE TO EUROPEAN COMMISSION CALL FOR EVIDENCE SURVEY ON "GENERAL PURPOSE AI CODE OF PRACTICE FOR PROVIDERS OF MODELS WITH SYSTEMIC RISK"

The Association for Computing Machinery (ACM) is the world's longest established professional society of individuals involved in all aspects of Computing. It annually bestows the ACM A.M. Turing Award, often popularly referred to as the "Nobel Prize of Computing." ACM's Europe Technology Policy Committee ("Europe TPC") is charged with and committed to providing sound **technical information** to policy makers and the general public in the service of sound public policymaking. Europe TPC has responded to the European Union stakeholder's consultations in the past in the context of the AI Act[1], the Data Act[2], the Digital Services Act[3][4], the Digital Citizen Principles[5], the Cyber Resilience Act[6], amongst others[7]. ACM and Europe TPC are non-profit, non-political, and non-lobbying organisations.

Europe TPC is pleased to respond to the European Commission's call for evidence launched on 13 November 2024 on the European Union's **"General Purpose AI Code of Practice for Providers of Models with Systemic Risks"**. Europe TPC supports the European Commission's intent on establishing key considerations for providers of general-purpose AI models and for providers of general-purpose AI models with systemic risk, through four Working Groups working in close collaboration with a pool of experts. Notwithstanding this general support, EuropeTPC would like to raise **twelve (12) recommendations** related to the code of practice as currently drafted.

Global Recommendations

- **Recommendation 1** - Although the document aims to introduce guidance for model *providers* with systemic risks only, a significant portion of the document (Measures 6+) seems to directly depend on the *deployment* and use-cases of the models in the context of encompassing AI systems. Many measures reference expectations for model *deployers*, but provide them under a Code of Practice that is targeted at model *providers*. The main recommendation from Europe TPC is that the code of practice should stay within the limits of its title "General Purpose AI Code of Practice

[1] https://www.acm.org/binaries/content/assets/public-policy/europe-tpc-comments-ai-consultation.pdf
[2] https://www.acm.org/binaries/content/assets/public-policy/acm-eur-tpc-data-act-comments-13may22a.pdf
[3] https://www.acm.org/binaries/content/assets/public-policy/europetpc-digital-services-act-comments.pdf
[4] https://www.acm.org/binaries/content/assets/public-policy/acm-europe-tpc-dsa-comments.pdf
[5] https://www.acm.org/binaries/content/assets/public-policy/europetpc-comments-digital-principles.pdf
[6] https://www.acm.org/binaries/content/assets/public-policy/acm-europe-tpc-cyber-reslience-comments-pdf
[7] https://www.acm.org/public-policy/public-policy-statements

for Model Providers with Systemic Risks". The Commission should therefore revisit the content of the Code of Practice to ensure that it encompasses solely the development and release stages of models by GenAI model providers, as opposed to encompassing both GPAI model providers and GPAI model deployers. Europe TPC recognizes that it may be challenging to separate the two scopes from a technical and domain perspective, hence an alternative would be to expand the scope of the Code of Practice to the deployment of models and to re-calibrate the participation in the working groups to ensure appropriate representation from model *deployer* organisations as well.

- **Recommendation 2** - EuropeTPC supports the commission's intent to define a code of practice for providers of models when introducing models with systemic risks, however it is important to recognize that a significant number of risks may arise in models *without* systemic risks that do not have the best practices in place, leading to potential systemic risks. EuropeTPC recommends that the commission makes clear that the processes and mitigations in place for models with systemic risks are a superset of the mechanisms that would be required for models with high and de-minimis risk. If possible the commission should provide guidance on which processes are a MUST/SHOULD/COULD for models with systemic, high and de-minimis risk respectively. Furthermore, such guidance should take into consideration the probability of the risk and the impact of the outcome(s)[8].

- **Recommendation 3** - The guidelines seem to be focused mainly on large-language foundation models, and large-image foundation models, however it is important to ensure that the code of practice applies to other modalities and associated systemic risks as well; examples include time-series, tabular, etc.

- **Recommendation 4 -** EuropeTPC would like to highlight that although a large percentage of the current discourse in context of "General Purpose AI Code of Practice" is in context of Generative AI models, the scope of General Purpose models can extend beyond purely on "generation of text, images and other content" as stated in II.b. Foundation models provided can also be made available in other modalities such as classification or regression interfaces.

Measure-specific recommendations

- **Recommendation 5 -** In context of *Measure 2*, Europe TPC suggests that the European Commission considers the standardisation of the metadata that is expected to be collected from model providers, and explores the possibility of providing templates to simplify the process of registering models and minimise overhead[9], particularly when multiple versions of a model may be released in a relatively agile manner. Furthermore it may also be important to consider the

[8] Any mitigations would need to be aligned with the probability and impact of the risk, encompassing a quadrant that can be suggested as part of the code of practice.
[9] These templates could encompass model-fair-use templates which are compliant with AI Act requirements, similar to how standardised licenses exist for open source code (e.g. Apache, MIT, etc)

monitoring to be put in place for model providers, such as enabling requests for correcting inaccurate information.

- **Recommendation 6 -** In context of *Measure 6*, the General-Purpose AI Code of Practice Draft highlights that one use-case that is treated as systemic risk is "Automated use of models for AI Research and Development". EuropeTPC would request that this is reconsidered and removed as a systemic risk, as the "increase the the pace of AI development" should not be seen as detrimental to the European scientific community and ecosystem. Furthermore, in *sub-measure 6.3.1. Dangerous model capabilities***,** Europe TPC points out "Long horizon planning, forecasting and strategising" is quite a generic term.. A significant subset of machine learning use-cases, such as demand forecasting, are leveraged quite commonly in industry for low-risk use-cases. The Code of Practice should better qualify when these capabilities should be considered a dangerous model capability[10].

- **Recommendation 7 -** In context of *Measure 10*, EuropeTPC recommends that when it comes to robust evaluation methods, these should at the very minimum ensure that relevant technical and non-technical domain experts have been involved to ensure fit-for-purpose evaluation. Furthermore, from a technical perspective, evaluations should be measurable[11], reproducible and auditable.

- **Recommendation 8 -** In context of *Sub-measure 10.3*, Europe TPC suggests that rather than purely seeking to establish a gold standard for operationalising high scientific rigour, the Commission identifies a set of parameters or metrics that can be examined for effectiveness on a regular basis. For example, such parameters or metrics could include (but not be limited to): verifiability of the reported findings, adherence to open data principles, declaration of competing interests, type of data used and its relevance to the scientific endeavour, correlation between the inputs and outputs, appropriateness and relevance of methods used, approaches to classification and clustering of data, justifications provided for research design, and roles and responsibilities of those engaged in the scientific endeavour.

- **Recommendation 9 -** In context of *Sub-measure 10.8*, Europe TPC recommends ensuring alignment and compatibility with independent organizations such as the UK AI Safety Institute[12] and the US AI Safety Institute[13], etc. as channels, organisations and methods that would facilitate the sharing of evaluations, tools and best

[10] The role of the Human In The Loop (HITL) should be a key consideration in any assessment of model capabilities deemed dangerous.
[11] To ensure measurability, these can be relative to specific business and/or impact thresholds, and/or relative to industry and academic benchmarks, for example. Specifically, one may suggest adapting to AI risk management the standard model employed in financial Operational Risk Management since 2000. The probability can be calculated by means of a set of consistent, robust and integrated compliance metrics inspired by the four principles in the AI act: (art 10-14-15): Sustainable, Accurate,Fair,Explainable (SAFE) . Existing SAFE AI metrics should be considered, such as https://doi.org/10.1016/j.eswa.2024.125239
[12] https://www.aisi.gov.uk/
[13] https://www.nist.gov/aisi

practices. These organisations can support the testing of high risk models, as well as defining standard frameworks that can be adopted to ensure robust and consistent testing.

- **Recommendation 10 -** In context of *Sub-measure 12.2*, Europe TPC would like to highlight a broad set of initiatives that are standardising taxonomies and risks within cybersecurity, such as the Institute for Ethical ML's MLSecOps framework[14], the MITRE Attack Framework[15], the OWASP Top 10 ML[16], and the UK National Cyber-security Centre Machine Learning Security Principles[17]; furthermore Europe TPC recommends the European Commission to set up respective initiatives to develop EU cybersecurity-specific standards by CEN, CENELEC or ETSI in support of legislation (AI Act) and the associated Code of Practice.

- **Recommendation 11 -** In context of *Measure 14*, EuropeTPC recommends that the term "deployment" is revisited to disambiguate it in this context, as it seems to refer in this section to "making the model available for use". However, the term "deployment" in the AI Act is used in the context of "model deployers" who integrate a model into a production AI system. Based on this, a more accurate term would be to "release the model", as the section specifically suggests mitigating the "release" of the models based on constraints provided. Once a model is released, model deployers would be able to deploy and integrate it into their AI systems.

- **Recommendation 12** - In context of *Sub-measure 18.2*, Europe TPC highlights that the consequences of any serious incidents need to be aligned with the type of capital (i.e. human, natural, social, manufactured, or financial) impacted. Any corrective measures identified need to be proportionate to the consequences of the serious incident on the impacted capital(s).

[14] https://ethical.institute/security.html
[15] https://attack.mitre.org/
[16] https://owasp.org/www-project-machine-learning-security-top-10/
[17] https://www.ncsc.gov.uk/collection/machine-learning-principles