Civil society's urgent warning:
The EU's Code of Practice for General Purpose AI final draft cannot abandon fundamental rights, and protections for children.

Dear Executive Vice-President Virkkunen,

We, the undersigned, are writing to express our concerns over the downgrading of fundamental rights, child sexual abuse material (CSAM) and nonconsensual intimate image (NCII) abuse risks in the third draft of the EU Code of Practice for General Purpose AI models ("the Code") to voluntary measures in an Appendix of the Code.[1] We believe the Code is in danger of underserving the regulatory framework set up by the AI Act for General Purpose AI models ("GPAI").

This letter acknowledges the Co-Chairs' efforts on the Code but expresses our collective concern that basic protections for fundamental rights and children remain unresolved at this late stage. Whilst every organisation and individual below is engaged in detailed argument across the Code, we are coming together to express our joint concerns, which we note are also reflected in a letter that was sent to you by the Representatives of the co-legislators, including the Spanish Presidency and the Rapporteur and Shadow Rapporteurs of the European Parliament for the European AI Act to the Commissioner published on 25th March.[2]

What are the consequences of the 3rd draft of the Code?

The description of risks as only "potential considerations" represents a radical change in the proposed Code's approach to fundamental rights, CSAM and non consensual intimate images. The consequences of this change include:

1) Changing the responsibility to assess risks from a process of risk investigation and discovery, based on principles set by the Code, to an optional risk selection process by model providers.
2) Shifting responsibility for assessments to downstream AI system providers, whereas the objective of the Code was to ensure accountability for upstream GPAI model providers;[3]
3) Undermining some potential benefits of independent external assessments, as these will be limited in scope to those risks identified by GPAI model providers;
4) Equally hindering the value of transparency requirements as GPAI model providers will be able to argue they do not need to share information with down stream EU based AI systems developers on optional risks they have not selected;
5) Rendering the Code unfit as a standards model, as fundamental rights are only included on an informational basis, rather than indicating where investigation should start to meet a standard;

---

[1] See '*Appendix 1.2 - Other types of risks for **potential** consideration in the selection of systemic risks* 3rd Draft of the EU Code of Practice for General-Purpose AI;

[2] MLex, *AI model providers see EU legislators raise 'great concern' on code of practice* 25 March 2025

[3] (Commitment II.11.1) only applies to systemic risks that model providers have themselves identified)

6) Diverging from international consensus such as that reached by the International Scientific Report on the Safety of Advanced AI. That report accepts fundamental rights risks as part of GPAI's capabilities and recommends methods to address them. The report states: "Several harms from general-purpose AI are already well established. These include scams, non-consensual intimate imagery (NCII) and child sexual abuse material (CSAM), model outputs that are biased against certain groups of people or certain opinions, reliability issues, and privacy violations…Since the publication of the Interim Report, new evidence of discrimination related to general-purpose AI systems has revealed more subtle forms of bias."

Addressing the arguments that have been proposed to support the 3rd draft

1) **Existing industry safety frameworks are not adequate to comply with the EU AI Act**

It has been argued that this[4] downgrade is to "make the AI Act compliance as easy as possible" by adopting a safety framework that "many companies are already using," but this misses the point. None of the industry safety frameworks mentioned in the explainer of the third draft, except one, set out to meaningfully address human rights and child sexual abuse content, and many, such as Meta and Google, omit the widely acknowledged risks of bias and discrimination from their frontier AI safety frameworks.[5]

2) **Fundamental and children's rights risks arise at model as well as system levels**

Another explanation provided has been that risks to fundamental and children's rights arise more clearly at system level, and so are better addressed by downstream system providers rather than at model level. It is the clear intention of the EU AI Act[6] to require providers to assess and mitigate all risks that are systemic, throughout the entire AI lifecycle, regardless of the existence of legal duties placed on others to assess and mitigate these same risks too.[6][7][8] [9] A case can be made that by leaving them out of the risks actually included within the framework, such as loss of control, the Code actually misleads GPAISR providers about the potential of their models to carry risks and their duties under the Act.

---

[4] See "Explainer about the Safety and security section of the Code," at the top of the code of practice published by the co-chairs on their website here: https://code-of-practice.ai/?section=safety-security#commitment-ii-3-systemic-risk-identification-2.

[5] See frameworks cited in the explainer here: https://assets.anthropic.com/m/24a47b00f10301cd/original/Anthropic-Responsible-Scaling-Policy-2024-10-15.pdf https://cdn.openai.com/openai-preparedness-framework-beta.pdf https://storage.googleapis.com/deepmind-media/DeepMind.com/Blog/updating-the-frontier-safety-framework/Frontier%20Safety%20Framework%202.0%20(1).pdf https://ai.meta.com/static-resource/meta-frontier-ai-framework/ https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/final/en-us/microsoft-brand/documents/Micro soft-Frontier-Governance-Framework.pdf https://images.nvidia.com/content/pdf/NVIDIA-Frontier-AI-Risk-Assessment.pdf https://cohere.com/security/the-cohere-secure-ai-frontier-model-framework-february-2025.pdf https://assets.amazon.science/a7/7c/8bdade5c4eda9168f3dee6434fff/pc-amazon-frontier-model-safety-framework-2-[6] -final-2-9.pdf https://x.ai/documents/2025.02.20-RMF-Draft.pdf Meta, Frontier AI Framework version 1.1;Google, Frontier Safety Framework 2.0.

[7] Recital 101 explicitly notes" Providers of general-purpose AI models have a particular role and responsibility along the AI value chain, as the models they provide may form the basis for a range of downstream systems".

[8] See Recitals 101, 114 and 115 of the AI Act.

[9] Engineering consortiums like MLCommons have created a "Hazard Taxonomy" for GPAI models, proving that practical hazard identification is possible https://arxiv.org/pdf/2503.05731

Finally, such an approach runs counter to the principle of risk management in line with other existing safety frameworks[10] which demonstrate that risks to fundamental rights are specifically linked to how GPAI models are pre-trained, set up and fine-tuned.

3) **Systemic risks include risks to public health, safety, public security, fundamental rights, or the society as a whole**

The final argument put forward stems from a particular interpretation of the risks unique to GPAI. According to this argument, the exclusion of fundamental rights risks from the "selected systemic risk" taxonomy, and inclusion under an optional category, follows an assumption that systemic risks must emerge as a result of high-impact capabilities.

That reading of the Act is flawed, as recently confirmed by the above-mentioned joint, cross-party letter of the representatives of the co-legislators. Indeed, it is difficult to reconcile how any interpretation of the definition of systemic risks pertinent to large-scale GPAI models which follows the AI Act's intentions would result in excluding assessment of fundamental rights, child sexual abuse material and nonconsensual digital abuse risks. Recitals 110-111 and Article 51 of the AI Act clearly mention that GPAI with systemic risks can arise from GPAI models with high-impact capabilities or an impact deemed equivalent. The co-legislators confirm their intent stating the systemic risks are inherently "significant due to their reach, or due to actual or reasonably foreseeable negative effects" and do not need further narrowing on severity.

In Conclusion

The Code of Practice has the power to sit firmly within the intention of the EU AI Act to set a global benchmark for human-centric and trustworthy AI governance. By enshrining protections for fundamental rights, democracy, and transparency, it can ensure that AI strengthens, rather than erodes, the freedoms that define open societies. The process has made many gains, but we fear that the chairs are under huge pressure to minimise the Code to gain maximum sign up. There has to be a limit to compromise. It lies in the faithful implementation of the EU AI Act.

We, alongside the representatives of the co-legislators, would have the gravest concerns over any version of a Code that treats the protection of human rights and crucial efforts to prevent child sexual abuse content as optional features. We urge you to ensure that the Code is fit for the purpose the AI Act intended and includes these risks in the list of types of systemic risk. Without these changes, the Code could legitimise a technology landscape where safeguarding fundamental rights and protecting children from exploitation are considered discretionary business decisions rather than moral imperatives.

[10] See US NIST, AI Risk Management Framework and US Department of State, Risk Management Profile for AI and Human Rights; Council of Europe, Framework Convention on AI, Human Rights, Democracy and the Rule of Law + HUDERIA Methodology; G7 Hiroshima AI Process; OECD, Towards a Common Reporting Frameworks for AI Incidents; UNESCO Recommendation on the Ethics of AI and related tools; 5Rights Children & AI Design Code.

Signatories

5Rights Foundation
Ada Lovelace Institute
Avaaz Foundation
ARTICLE 19
Centre for Democracy and Technology Europe
Dr Chiara Gallese - Eindhoven Artificial Intelligence Systems Institute
Digihumanism - Centre for AI and Digital Humanism
Dr. Karine Caunes, Associate Researcher, Lyon 3 University; Editor-in-Chief, European Law Journal
Dr. Marta Bieńkiewicz, Z-Inspection - Trustworthy AI Initiative
David Evan Harris, University of California, Berkeley
European Center for Not for profit Law
European Writers' Council (EWC) - Nina George
Piotr Brzyski Zetatech
Pour Demain
SAIL - Stewardship AI Lab - Kroplewski Law Office - Robert Kroplewski,
The Future Society - Toni Lorente