Brussels, 25 October 2022

Call for better protections of people affected at the source of the AI value chain

Dear Members of the European Parliament,

We, a group of leading non-profit and civil society organisations, call for **ex-ante obligations in the AI Act on the providers of general purpose AI systems (GPAIS).**

GPAIS are AI systems that can accomplish or be adapted to accomplish a range of distinct tasks, including some for which they were not intentionally and specifically trained[i]. The last years have seen a surge in such systems across a spectrum of abilities such as language, vision, robotics, interaction, understanding, reasoning, and search. GPAIS include both unimodal (e.g., GPT3 and BLOOM) and multimodal (e.g., Stable Diffusion and Dall-E) systems[ii] and they can be trained through different methods – whereas e.g. Gato uses supervised learning, MuZero is based on reinforcement learning.

The trend towards more general and capable systems is unmistakable and these systems come with **great potential for harms** if left unchecked. GPAIS have already caused alarm by propagating extremist content[iii], encouraging self-harm[iv], exhibiting anti-Muslim bias[v], or inadvertently revealing personal data[vi] among many other harms.

While general purpose AI systems were not part of the Commission's proposal, both the European Parliament and the Council proposed specific provisions to address them explicitly in the AI Act. Although the co-legislators have recognised the need to regulate GPAIS, this is not enough. What matters is how these systems are regulated to ensure that they are **safe** when placed on the market and that they **protect people's rights**. Adequate regulation of GPAIS will also ensure **legal certainty** and close loopholes.

In this context, it is crucial that the responsibility to comply with the obligations of the AI Act be shared between the providers (developers) and the users (deployers) according to their level of control, resources and capabilities.

There are only a handful of providers of GPAIS who are all very well-resourced with huge computational capabilities and who employ the world's best AI researchers. A single GPAIS can be used as the **foundation for several hundred applied models** (e.g. chatbots, ad generation, decision assistants, spambots, translation, etc.) and any failure present in the foundation will be present in the downstream uses.

Therefore, providers of general purpose AI systems should fulfill all requirements in Title III, Chapter 2, such as set up risk and quality management systems, put in place data governance practices, draw up technical documentation, and test the accuracy and robustness of their systems. These obligations should apply regardless of the means by which these systems are later made available to downstream users.

Shifting these obligations to downstream users would make these systems less safe because these users will not have the same capacity to change or influence the behaviour of the model. Users will generally not have control over the model. Even if they did, they are unlikely to have the capacity to process and understand the vast amount of model data. The downstream users are also not aware of the design logic used by the upstream providers. These design choices can result in safer or less safe systems.

However, this does not mean that users should be off the hook. **Users are best placed to comply with requirements in relation to the specific high-risk use case**. These include human oversight, but also any use-case specific quality management process, technical documentation, logging as well as any additional robustness and accuracy testing. These requirements should apply to the users especially when the use cases are novel and cannot be reasonably foreseen by the providers.

Concluding, the AI Act must place specific ex-ante obligations on the providers of general purpose AI systems, not only to ensure an adequate allocation of responsibilities across the value chain but also to protect people from manipulative, biased or discriminatory systems, which could have devastating effects on them.

We hope you take the necessary steps to address our concerns and we are looking forward to further engaging with you to ensure a better protection of people affected at the source of the value chain.

Sincerely,

Access Now

Bits of Freedom

Electronic Frontier Norway (EFN)

European Center for Not for Profit Law (ECNL)

European Digital Rights (EDRi)

Future of Life Institute

Homo Digitalis

Irish Council for Civil Liberties (ICCL)

Panoptykon

The Future Society

[i] Gutierrez, Carlos Ignacio and Gutierrez, Carlos Ignacio and Aguirre, Anthony and Uuk, Risto and Boine, Claire C. and Franklin, Matija, A Proposal for a Definition of General Purpose Artificial Intelligence Systems (October 5, 2022). Available at SSRN:https://ssrn.com/abstract=4238951.

[ii] Unimodal systems have one core ability such as text processing; multimodal systems include combinations such as text processing plus image generation. Unimodal systemscan be general-purpose if they accomplish a range of distinct tasks.

[iii] https://www.middlebury.edu/institute/sites/www.middlebury.edu.institute/files/2020-09/gpt3-article.pdf.

[iv] https://www.nabla.com/blog/gpt-3/.

[v] https://arxiv.org/abs/2101.05783.

[vi] https://www.theregister.com/2021/04/29/scatter_lab_fined_for_lewd_chatbot/.