Open letter: AI needs rules, not pauses

Why we, as organizations and individual citizens, believe that the proposed moratorium and much of the current debate around generative AI systems is not enough to protect our democratic society and is missing the point on how to ensure safeguards and fundamental rights.

Milan, 03/04/2023

Dear members of the European and Italian Parliament,

As a group of organizations and individual citizens engaged in the defense of civil rights and the promotion of ethical and equitable technologies, wish to draw your attention to our concerns regarding the development and implementation of artificial intelligence (AI) systems, including *Large Language Models* (LLMs) and generative models. **We are deeply concerned** by the implications of these technologies for our society and for the future not only of Europe, and we believe it is crucial to address these challenges proactively, collaboratively, and responsibly.

We address this letter primarily to the European Parliament, which is due to vote on the political agreement on the *Artificial Intelligence Regulation* (AI Act) on April 26, but also to the Council of the EU and the European Commission, which will participate in the trilogues from May, and to all civil society interested in balancing the benefits of AI by addressing the issues related to its impact on society, people, and the environment.

In recent weeks, the focus on the developments of generative AI has led to an intense and complex debate on the regulation and control of these technologies. Models such as *GPT-4*, *DALL-E*, *Stable Diffusion*, and *AlphaCode* are becoming the basis of many computing systems in diverse contexts, such as sales, customer service, software development, design, gaming, education, and many others. However, the use of these generative models carries with it risks and potentially problematic implications, such as **the spread of disinformation, the generation of prejudiced and stereotypical content, the emulation of extremist content, and the creation of increasingly realistic deepfakes**.

We acknowledge the importance of considering geopolitical issues and international collaboration in addressing these problems and commit ourselves to working with you to promote a balanced and sustainable approach to the use of AI not only in the latest applications and those that will continue to develop, but also in creating a future-proof European regulation focused on the preventive assessment of compliance with the protection of fundamental rights and safety principles.

On the *Future of Life Institute*'s Moratorium

Accordingly, we would like to draw your attention to the recent moratorium proposal put forth by the *Future of Life Institute* on March 28, 2023. This moratorium is an open letter that has so far collected over 1,800 signatures from experts and ordinary citizens alike. The proposal calls for a six-month suspension in the development of generative models such as current LLMs, defined in an overly ambiguous manner as "*powerful AI systems with human-competitive intelligence.*" The goal of this moratorium is to demand that generative AI development be focused on making "*powerful, state-of-the-art systems more accurate, safe, interpretable, transparent, robust, aligned, trustworthy, and loyal.*"

While some of the concerns expressed in the letter are shared, we believe that it presents certain **limitations and issues that are diverting attention from what truly requires commitment**. Public opinion is rightly focusing on these models, but we think that the inflated and exclusive hype surrounding these applications is distracting us from the concrete impacts and the ways to mitigate them. In the following points of this

letter, we will outline our concerns regarding the developing narrative and the aspects of the proposed moratorium that we consider insufficient or problematic:

- The substantial invalidity of any appeal and/or time frame proposed to suspend the development of such systems, given the outdated solutions contained therein at the **geopolitical and market reasons**, as well as the **dangerous enabling of a narrative** on AI that portrays it as something that will surpass us in human cognitive abilities and render us obsolete;
- The need to address more concrete and pressing issues related to the ecosystem of generative artificial intelligence technologies, such as the **growing concentration of power** in the hands of technology companies, the challenges for adequate regulation, and issues related to digital colonialism;
- The importance of a **proactive approach to risk management** and the **preventive assessment** of these technologies that involve a wide range of stakeholders in the decision-making process;
- The promotion of **transparency at all stages of the AI development and usage** process in order to ensure an effective approach to risk management and the control of technologies;
- The importance of considering the potential **psychological harms** caused by generative AI and developing strategies to prevent and mitigate such negative effects on individuals' mental health;
- The need for **international cooperation** and the **development of shared regulations and policies** at the global level to address the risks and challenges of generative AI in a harmonized manner.

We believe these issues are fundamental to ensuring a balanced and sustainable approach to the use of generative AI and promoting a safer, more ethical, and equitable future for all. We invite you to take these concerns into consideration in your assessment of the proposed moratorium and in your commitment to the regulation and oversight of such technologies.

Perspectives, issues, and possible solutions

Starting from the moratorium, the narrative used to represent the development of these AI systems seems to legitimize a narrative of AI that will soon be conscious, thus

capable of exercising power and posing a threat to humanity as a whole. These sci-fi narratives **shift attention away from the already problematic risks and impacts, such as discrimination or exclusion of certain social groups, and prevent the identification of responsibilities**. It is crucial to consider how states, their policies, and power dynamics among them significantly influence the way artificial intelligence models are developed and used. Global competitiveness in the field of AI and the strategic interest of the world's major powers in such technology make it essential to address geopolitical issues to ensure fair and ethical AI development. Such issues include:

1. ***Regulatory and value divergences*** can hinder the creation of a shared regulatory framework at the global level, but at the same time must be acknowledged as fundamental in the self-determination of cultures.

2. ***Risk management, assessment, and transparency in AI*** play a fundamental role in ensuring accountability and trust in the AI development and usage process. It is therefore essential to promote the adoption of tools and practices that enable effective risk assessment and management associated with these technologies. For this reason, we urge you to:
   a. Develop tools and methodologies that allow for the early identification and quantification of AI-related risks in terms of potential negative impacts on **fundamental rights, society, safety, security, and the environment** directly addressing emissions related to AI development, requiring periodic reviews.
   b. Auditing of such systems, carried out by independent and qualified entities, helps ensure compliance with existing regulations, clear documentation of development methodologies, learning models, and evaluation criteria used, as well as the ability to examine and verify underlying decision-making processes.
   c. Support research and development of tools and methodologies for risk management and auditing in AI, to ensure safety, reliability, and compliance **even when not considered high-risk**. It would be functional to **expand transparency and publicity obligations to all AI systems in use by "public" entities** regardless of the level of risk. The enrollment to the European database of all AI systems in use by the public, regardless of the degree of risk.
   d. Promote the training of specialized personnel in the analysis and management of risks associated with AI, to ensure competent and independent evaluation of AI systems. In this perspective, the newly established *European Centre for*

Algorithmic Transparency (ECAT) represents a direction that is as favorable as possible to these points.

3. ***Ecosystem of regulations:***
   a. The EU *Digital Services Act* (DSA) represents an important step towards the regulation of generative AI, encouraging the involvement of the online platform communities in the process of monitoring and evaluating AI technologies.
   b. The EU *GDPR,* as witnessed by the recent blockage of ChatGPT on a national level by the Italian *Data Protection Authority*, is a complementary legal baseline for data treatment that should constitute, among the upcoming *Data Act* and *Data Governance Act,* the regulatory benchmark where generative AI providers should proactively inform their data treatment policies.

4. ***Asymmetries in the distribution of resources and skills***: countries with greater technological and financial capacity can influence the development and use of AI at the expense of less advanced countries, increasing inequalities and the risk of digital colonialism. To address these challenges, we urge the European Union to:
   a. Discourage competitiveness and overpowering mechanisms in the event of temporary suspensions, encouraging the creation of collaboration networks between research institutions, technology companies, and governments, in order to mitigate the effects of geopolitical rivalries on AI cooperation.

5. ***Digital colonialism***: this phenomenon ranges from the extraction to the appropriation of data and digital resources by more advanced countries and companies at the expense of developing countries. The main consequences and ethical issues of digital colonialism are (i) **human exploitation** - as in the case of Kenyan workers exploited by OpenAI at two dollars an hour to train their AI; (ii) **economic exploitation** - the monopolization of digital resources on the one hand and the inability to develop a local industry; (iii) the creation of **technological dependencies** - the dominance of a single model of digital infrastructure creates dependencies and prevents the development of local architectures. To counter this phenomenon, the European Union should:
   a. Promote fair and transparent regulation for data collection and use, ensuring respect for the privacy and human rights of the populations involved;
   b. Support international cooperation initiatives that promote the transfer of technologies and skills to developing countries, allowing an autonomous contribution to the generative AI offerings;

c. Encourage the creation of accessible and democratic digital infrastructures and artificial intelligence services, which can guarantee greater control and autonomy for less advanced countries.

d. Power dynamics and concentration of control over AI results in an almost monopolistic control over the development, dissemination, and implementation of AI technologies. As a result, decisions concerning ethical standards, research priorities, and practical applications of AI tend to be driven by the interests of these entities rather than by global consensus or common interest.

6. ***Reducing the overall impacts of AI on the climate*** and defining impact requirements for **emissions associated with the development and use of AI**. To act responsibly within the context of climate change, it is important to avoid adopting a "*techno-solutionism*" approach and not to overestimate the ecological costs of AI. Therefore, it is essential to align the entire AI ecosystem with environmental sustainability strategies. Governments should adopt regulations, strategies, funding, and AI procurement programs that **take into account the climate impact**, e.g. avoiding government funding of applications that are contrary to climate objectives and purchase AI services only from companies committed to a zero-emissions goal;

a. Ensure the adequate inclusion of cloud computing in reporting and carbon pricing policies;

b. Establish reporting requirements and data availability for the life cycle emissions of AI;

c. Define methodological standards for environment impact assessments at the national and international level.

7. ***International cooperation for AI governance among industry players, regulators, and civil society.*** The pervasive and cross-border nature of AI requires a coordinated effort so that regulations and policies adopted globally are as aligned and complementary as possible. This can contribute to the sharing of best practices among countries, facilitating the adoption of common standards and the creation of a shared regulatory framework that takes into account the cultural, economic and social specificities of different regions of the world.

a. Through participation in international initiatives, fora, and collaboration with other international organizations to promote the adoption of common principles and guidelines for AI globally.

b. By planning dedicated task forces including citizens, civil society and relevant stakeholders to discuss and identify possible AI's societal pitfalls and elaborate on specific democratic and shared solutions.

c. In addition, there are also other reporting mechanisms that allow users to report false, biased, or harmful results created by AI technologies, such as Facebook's reporting mechanism or Google's misinformation reporting platform. However, these mechanisms need to be further developed and enhanced, as within the upcoming EU *Digital Services Act* to ensure greater effectiveness in managing the risks and challenges with generative AI.

Conclusion

Europe has the opportunity to take on a leadership role in promoting a democratic and inclusive approach to AI development, ensuring fairness and justice in the distribution of benefits offered by these technologies. Promoting transparency and adequate risk management in generative AI is crucial to ensure the security, stability, and trust of citizens in artificial intelligence technologies.

In particular, we believe it is necessary to address issues such as **digital colonialism and the concentration of control** over AI in the hands of a few technology companies and nations to ensure fairness and justice in the distribution of generative AI benefits. Furthermore, **promoting a proactive approach to risk management**, involving a wide range of stakeholders in decision-making, is essential to prevent potential negative effects resulting from the use of generative AI.

To achieve these goals, the European Union has the potential to lead the development of a regulatory framework and innovative practices that focus on responsibility and ethics in the age of artificial intelligence. International cooperation in AI governance is essential to ensure a balanced, ethical, and sustainable approach to the development and use of artificial intelligence technologies.

Therefore, we appeal to Members of the European Parliament to promote a collaborative and inclusive approach to generative AI governance and to ensure that artificial intelligence technologies are used for the common good.