

EARE's response to EU Commission's consultation on trustworthy general-purpose AI models under the Artificial Intelligence Act

The European Alliance for Research Excellence (EARE) welcomes the opportunity to express views on the topics covered by the first Code of Practice which will detail out rules for general-purpose Artificial Intelligence (GPAI) models. The AI Office has a critical role to play in shaping the future of AI, and the development of this first Code of Practice is crucial for the future of GPAI models in Europe. Startups, researchers and research performing institutions are core to the economic growth of the European Union. Any Code of Practice that impacts research activity should ensure that research activity and translational research is not overlooked. This Code should strike the right balance, enabling researchers, startups, and innovators to fully harness GPAI models' potential, drive research excellence, stimulate innovation, and contribute to closing the EU's innovation gap.

Articles 3 & 4 of the Copyright Directive (2019/790) should not be circumvented

Today, research activities in Europe rarely benefit from the Text and Data Mining (TDM) exception outlined in article 3 of the Copyright Directive, because the majority of European research is done through public private partnerships or benefits from private funding. As such, when developing the template for the "*sufficiently detailed summary of the content used to train the general-purpose AI*" the AI Office should consider that it is in fact the requirements under Article 4 of the Copyright Directive that will be applicable to many areas of research. Public Private Partnerships, and the translation of research into the commercial sphere should be a key consideration when drafting this first code of practice.

In addition, Article 3 prevents contractual provisions from overriding the TDM exception, yet, libraries are observing that commercial providers are seeking to prevent the act of data mining for any purpose within contractual agreements in conflict with EU copyright and AI laws. The future code should further enforce the legitimate exercise of exceptions for data mining – including the ability to data mine for the purpose of training AI models – and it should prohibit any attempts to restrict such activities.

Transparency should empower, not hinder, innovation and research excellence

Transparency is a core value of the EARE. Yet, EARE believes that the current focus on scrutinizing the data and methodologies used to train AI models, while important, may inadvertently place excessive constraints on innovation. As recently mentioned in European Commission's [study](#) on improving access to reuse of research results for scientific purposes, a detailed summary of the data used for training can "*add a layer of compliance costs for research organisations*". Instead, the EU could benefit from shifting its focus towards ensuring the development of high-quality AI systems within the EU, especially given that the EU already benefits from existing legal frameworks that address data protection and rights.

When it comes to the template for the summary of training data used to develop GPAI models, EARE considers **this summary should be simple and should strike a balance**

between detail and clarity. First, such a document can be overly complex and difficult to manage, especially for SMEs, research organisations, and startups. Second, the summary should not request the disclosure of sensitive data, such as trade secrets. Protecting trade secrets is crucial for startups to secure funding, so any attempts to reveal this information should be avoided. Finally, on a very practical view, a high-level summary would ensure the information is accessible to a broad audience, including non-experts, contributing to the objective of transparency and trust and would facilitate the implementation of the AI Act requirements.

When developing the transparency requirements for GPAI models, **the Code of Practice should be crafted in a way that enhances the ability to innovate and drive research excellence in the European Union.** The AI Office should focus on essential aspects, ensuring that the requirements are clear, avoiding overwhelming detail and providing flexibility to adapt compliance efforts based on specific needs and circumstances. The AI Office should also learn from the UK experience where the UK Intellectual Property Office and the government were not [able](#) to produce a code that was balanced for all stakeholders. The future code of conduct for GPAI models should be properly balanced, and consider the views of researchers, startups, and innovators. The European Commission should also regularly review and update this code based on technological advancements and in collaboration with researchers, startups, academia, and industry.

Focus on copyright compliance, not overreach

Regarding the obligation to put in place a policy to comply with Union copyright law, EARE believes that the EU's Copyright Directive already provides sufficient safeguards by giving rightsholders the option to reserve the right for their works to be used to train algorithms. Rather than introducing new legal provisions without a comprehensive impact assessment to guarantee proportionality, the European Commission should focus on harmonising the implementation of the Copyright Directive across the EU single market, and fostering knowledge sharing and valorisation. The AI Office should develop recommendations and best practices that do not exceed the scope of the Copyright Directive to not lead to unexpected consequences, such as regulatory conflicts, increased administrative burdens, or adverse effect on businesses and research.

Rather than focusing on restricting data and creating burdensome requirements for AI providers, it is crucial the Commission ensures a large access to high-quality data. Text and Data Mining (TDM) is fundamental for training machine learning models and research. High quality and diverse data mined through TDM, and in line with EU regulations, helps in enriching AI models. This is essential not to undermine research, as restrictions increase the chance of bias in models. To stay competitive, the EU should consider ways to make more data available to enable AI applications and promote data and knowledge sharing. While reservation of rights and opt-outs are important, they are preserved in the Copyright Directive. They also present challenges for research capacities by limiting data access. **Efforts to promote data sharing**

and TDM practices and improved access to data should be the priority to foster a more conducive environment for research and innovation.

Conclusion

Our alliance welcomes the opportunity to contribute to the drafting of the Code of Practice for GPAI models. Our alliance supports the aim to consolidate the existing transparency requirements established in the AI Act. However, the AI Office should steer clear of overly restrictive measures and should not impose further obligations which can hinder innovation and research. The information provided by GPAI models should be comprehensive, simple, and include practical information, avoiding technical details, with the aim to facilitate researchers' work and avoid unnecessary burdens on European startups and innovators.

Only a balanced approach can empower researchers, startups, and innovators to unleash the full potential of GPAI models, driving research excellence, fostering innovation, and helping close the EU's innovation gap. Echoing Mario Draghi's report and its recommendations, it is crucial to consider AI development as an opportunity for European industry players to boost their competitiveness. Instead of focusing on restrictions, we should ensure open sharing of data and knowledge to help Europe deliver on its economic, social, environmental and democratic potential.

About EARE: *The European Alliance for Research Excellence (EARE) was convened in 2017, and now brings together seven members from the research and innovation ecosystem in Europe, including BSA | The Software Alliance, Allied for Startups, LIBER, LACA, Research Libraries UK, SCONUL (Society of College, National and University Libraries) and UCL (University College London) Library, advocating for the EU to live up to its innovation potential in the digital economy.*