



## European Alliance for Research Excellence

### ICO Consultation – Data Protection and Generative AI

The [European Alliance for Research Excellence](#) (EARE) is a coalition of companies and research organizations formed in 2017 that are committed to the future of innovation and R&D in Europe. We bring together universities, large and small technology companies, startups and scaleups, libraries, scientific and research funding and performing organizations, all united by the idea that access and use of data should be *“as open as possible, and as closed as necessary”*. Our members are either users, suppliers or developers of Generative Artificial Intelligence models, and therefore we wish to provide our collective input to this public consultation.

The EARE strongly believes that the development of AI should not be hampered, when there are ways to comply with GDPR requirements, including existing exemptions for scientific research purposes, allowing Generative AI developers and suppliers to train their models on datasets that respect privacy.

Generative AI models are trained on data that is collected by web scraping that may include personal information. However, what makes Generative AI performant and effective is the quality and quantity of training data, particularly the inclusion of rich and diverse datasets, which may contain personal information. This use of personal data and reliance can present privacy issues. EARE members believe that effective data research and innovation, including the training of AI models with data that includes personal data can comply with GDPR requirements, without having to constrain the use of downstream models. When it comes to developing and training Generative AI models, EARE members wish to underline that the collection and analysis of data, including AI training is vitally important and is in the public interest, providing societal and economic benefit. Steps can be taken to ensure that an individual's interests are balanced, thereby ensuring compliance with GDPR. These steps could include ensuring that technical measures to restrict access to data is respected.

Technical solutions such as anonymization and pseudonymization of data exist and can be used more broadly to guarantee that privacy is respected whenever new training datasets are created. Beyond this, ensuring compliance with GDPR can be intricate as it requires a deep understanding of how the algorithms works and how legal requirement outlined in GDPR need to be applied, but several measures can help mitigating the risk and ensuring data protection.

Respecting privacy and adhering to data protection laws are not only ethical imperatives but also essential for legal compliance, building trust with users, maintaining security, and fostering future innovation in the field of Generative AI.

Thursday, March 7, 2024

### Implementing transparency measures

Transparency and reproducibility are key principles for EARE. In the context of web scraping, this entails ensuring users who share personal information that may be made available online are well-informed about how their data may be made available. Empowering users with control over their data through tools like privacy dashboards and offering comprehensive transparent privacy policies online should be generalised. Furthermore, materials, such as FAQs and documents outlining privacy practices, aim to educate users about AI design, testing, deployment, and address ethical considerations like fairness, privacy, security, and accountability.

### Security and privacy measures

Data breaches present high risks for data protection. The risk that generated output might contain sensitive information or inadvertently personal data, for instance, following data breaches exists but can be mitigated with advanced IT security measures. Implementing technical and organization measures to ensure data is secured can significantly mitigate those risks, to the point that this almost meaningless.

### Ensuring compliance with current legislations

EARE supports collaboration with Governments, regulators, and other stakeholders to ensure compliance with current laws. We note however that the ICO guidance refers to intellectual property law. The training of AI on copyrighted works is not a copyright infringement and the ICO guidance should not imply that it is.

We welcome this consultation as we believe that Generative AI developers have an important role to play in terms of education and on how to adapt to this evolving technology. Privacy is a fundamental right and collaboration is key to ensure harmonized regulations which do not hinder innovation. Compliance with current legislation requires a complex and multidisciplinary approach, involving expertise in AI, ethics, law and policy that can be very burdensome, especially for startups, SMEs or researchers. This is even more complex when we are dealing with downstream providers or open-source models. A balance between openness and collaboration to protect individual privacy and data rights will have to be found, without hindering innovation and collaboration.