

data.europa academy webinar ‘Next steps to compliance: preparing for the EU AI Act’ 18/07/2025 - Q&A

Question	Answer
Please advise if a guideline for general purpose AI apply and governance is available?	Guidelines on the scope of obligations for providers of general-purpose AI models under the AI Act Shaping Europe's digital future
Could you share with us an acceptable organizational model structure in a private enterprise for handling AI topics?	There is no one acceptable organizational model structure for handling AI topics, it depends on your activities, your sector, the size of the company, your team and the use case portfolio, etc. Following the regulation and the international standards like the ISO24001 will help you to setup the right .
What role will the EU Data Union Strategy play?	The EU Data Union Strategy aims to establish a single market for data, enabling data to flow seamlessly across sectors and borders under clear rules, strong governance, and robust protections. At the heart of this vision is open data, which serves as a key enabler by providing accessible, high-quality datasets that fuel innovation. Artificial intelligence stands to benefit significantly from this ecosystem, gaining access to diverse, interoperable, and trustworthy data sources essential for developing impactful and responsible AI systems.
How does Ines see the interplay between the EU AI Act's pressure to use transparent datasets and the confidentiality exception for AI model providers (mentioned in the new GPAI Code of Practice)?	I believe the balance between transparency and confidentiality in both the EU AI Act and the GPAI Code of Practice is fair and pragmatic. The EU AI Act rightly emphasizes the need for transparency around datasets used to train foundation models, particularly to ensure safety, traceability, and non-discrimination. At the same time, the GPAI Code acknowledges the commercial and security concerns model providers may face by introducing limited confidentiality exceptions. This approach serves the broader ecosystem well. It creates a baseline of accountability and shared understanding while still protecting innovation. I'm particularly interested in the value this will unlock across the value chain, from developers to downstream providers. With more structured disclosures and documentation, we'll likely see

	<p>fairer and more tangible benchmarking between models, which can drive better choices, compliance by design, and ultimately, more trust in GPAI systems.</p>
<p>Is there a collaboration from INSEE in France in the Health Data Hub?</p>	<p>Yes, INSEE has been involved in collaboration with the Health Data Hub (HDH) in France, as a trusted “data-matching partner”, providing the infrastructure and methodology for pseudonymization and secure linkage required to combine data.</p>
<p>I support academic researchers daily in Research Data Management, and they are pressured to publish Open Research Data, but also want to be cited: how to enable FAIR datasets to be automatically cited when embedded in LLMs or used as RAG or sources for AI?</p>	<p>To enable FAIR datasets to be automatically cited in LLMs or RAG systems, datasets must include machine-readable metadata with citation details (e.g. author, license). This metadata should be preserved throughout the AI pipeline. In RAG, the retriever can return both the content and its source, allowing the LLM to generate responses with proper attribution.</p>
<p>How does Ines think about the current fair use of copyright litigation going on in the US and UK in terms of their relevance for open data?</p>	<p>The ongoing copyright cases in the US and UK, like <i>New York Times v. OpenAI</i>, reflect a growing tension between innovation and IP rights. While focused on proprietary content, their outcomes will influence global norms, including around open data.</p> <p>These cases show that even open data must be used carefully. Fair use is still unclear in the AI context and varies by jurisdiction. The US offers more flexibility for text and data mining, but that’s now being challenged. In contrast, the EU takes a stricter stance, requiring explicit authorization unless a legal exception applies.</p> <p>In short, open data is not a free pass. Developers must check licenses, provenance, and usage rights. As scrutiny grows, transparency and traceability in AI training data are key to legal certainty and public trust.</p>