Kluwer Copyright Blog

Copyright, Compliance, and Confidentiality: Finding Common Ground in Generative AI

Enrico Bonadio, Eduardo Alonso (City St. George's, University of London) and Vansh Tayal (Symbiosis Law School, Pune, India) · Wednesday, June 11th, 2025

Image by Tung Nguyen from Pixabay

The rise of generative AI and automated content generation has raised legal and ethical issues, making them a focal point in creative and technological sectors. As stakeholders navigate this new terrain, the EU AI Act appears as a benchmark regulatory framework. This blog briefly examines the transparency provisions and trade secret protection under the Act, not as sterile concepts of law but as relevant issues that affect interactions between creators, developers, and regulators in generative AI.

By framing the discussion through such practical lenses as artistic integrity, market competition, and societal trust, actionable insights are provided for legal practitioners and technologists. We here concisely highlight the opportunities and challenges these provisions present, especially vis-à-vis the balancing of copyright protections and innovation. Ultimately, the piece intends to help rights holders and AI developers proactively confront potential legal friction by encouraging meaningful discourse and steering the future of generative AI to honour the twin pillars of creative expression and technological progress.

The Opacity Challenge and the EU AI Act: Transparency vs. Trade Secrets

It is commonly noted that generative AI models typically act as a mysterious black box. The catch is that a person can observe the inputs and outputs, but not the main reasoning behind them. This

opacity makes it strenuous for copyright owners to know if, when or how their works were ingested or somewhat used without permission for AI training purposes.

The EU AI Act has taken a pioneering step by requiring providers of general-purpose AI models to provide sufficiently detailed summaries regarding the data upon which their training was carried out. At the same time, AI companies worry that such transparency obligations would be expensive and expose proprietary architectures and trade secrets, jeopardising innovation and competitiveness.

The relevant provision of the EU AI Act is Article 53(1)(d), requiring providers of general-purpose AI to prepare and publish a 'sufficiently detailed summary' of the dataset used for training, according to a template established by the AI Office. The dataset summary, provided in Annex XII, must contain at least information on content categories, sources, volume estimation, and methods of processing involved, but does not entail sharing raw data files.

Copyright holders view these summaries as a pivotal tool for discovering unauthorised usages and requesting compensation. Yet, as mentioned, AI developers argue that a wide blanket disclosure would disclose crucial aspects of model architecture, training methodology and proprietary pre-processing pipelines; and small and medium-sized enterprises fear that the cost and complexity of preparing these summaries might even prevent their capacity for innovation. Although the EU AI Act provides some SME-friendly measures facilitating access, priority fee-exempt participation to regulatory sandbox, simplified documentation templates, and proportionate assessment fees, this strict set of rules might entail increased costs, stronger bureaucracy, and hurdles at market entry, and consequently limit the involvement of smaller players in the AI ecosystem.

To alleviate these pressures, Recital 107 of the EU AI Act allows companies to benefit from a trade-secret defence for withholding pipeline or algorithmic details provided that the summary is generally comprehensive in its scope instead of technically detailed. The trade-secret exemption has been claimed by industry groups to be paramount in protecting innovation, posturing that divulging too much detail will hand competitors 'the keys to the kingdom.'

In contrast, full-transparency advocates caution that in the absence of well-defined, case-by-case limitations and independent review, the trade-secret defence could devolve into a 'blanket excuse' for intransparency, thus undermining the very goals of accountability enshrined in the Act. Indeed, several vendors of AI-based products exploit the narrative of 'inherently unexplainable AI' to justify very few or minimal disclosures. **In response to this self-serving stance**, the European Data Protection Board has asserted that black-box opacity cannot constitute a valid ground for avoiding transparency obligations either under the GDPR or under the EU AI Act. This could imply that certain platforms may be taking unscrupulous advantage of the claim of inscrutability to exercise delay or to circumvent their reporting obligations. This is also why, it has been argued, secrecy claims must be strictly circumscribed and justified by way of a public interest test.

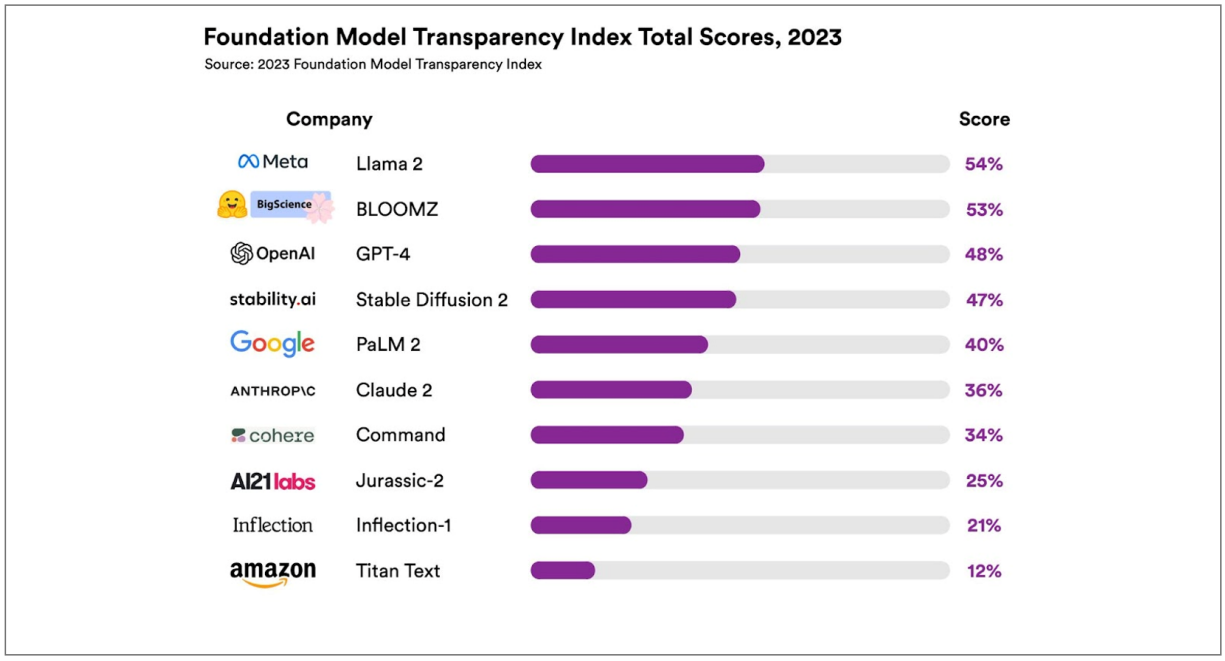
Practical Compliance Strategies

Since this tension exists, it is important to come up with templates for transparency that protect real trade secrets while furthering the EU AI Act's intention to empower rights-holders and build trust. Given this need to reach a balance between transparency and confidentiality, what practical steps can platforms take to meet the Act's mandates? First, in accordance with the EU AI Act, Article

53(1)(d), these entities may produce high-level summaries of training data that mention general classes of sources without revealing proprietary details. According to Recital 107, dataset providers may employ broad-type descriptions of the datasets; for instance, web scraping social media posts or licensing books and articles to avoid specifying each URL, individual title or database entry.

The exclusion of proprietary model details is an intentional feature of the EU template. It requests information on data sources and processing but does not request algorithms and architectures or exact pre-processing methods. For example, a company might record, 'The model was trained on ~10B tokens from public news sites (2015–2023), Wikipedia dumps, and a cleaned subset of social media text.' Such a statement would be sufficiently detailed and thus would enable rights holders to understand the broad scope of the content while protecting the company's investments in the content. Put another way, emphasising just the source types and the general methods used in preparation affords perfect compliance without giving away trade secrets.

For example, the Foundation Model Transparency Index developed by researchers at Stanford University evaluates whether a major model really fulfils the criterion: Meta's Llama 2 scores just 54 percent, and OpenAI's GPT-4, a mere 48 percent, on 100 transparency metrics. The very low grades can somehow indicate the evident gap that exists between legal obligations and actual behaviour, advocating for more viable measures of transparency.



Source: Stanford University

Beyond dataset summaries**,** another approach could be to adopt standardised model cards as a pragmatic middle ground. First proposed in 2018 by Mitchell et al., model cards accompany trained AI models with documentation that explains their intended use cases, categories of training corpus, metrics used in measuring performance, and any known shortcomings. The cards promote transparency by divulging high-level information regarding data sources, without the card issuer revealing any details deemed proprietary. Organisations like Google have developed model cards on their AI platforms to help end users assess model suitability. Privacy and governance bodies

such as IAPP recommend model cards to promote responsible AI deployment. AI providers may want to extend this approach with a dedicated copyright-transparency section, broadly listing data sources, such as license-holder publisher archives, public domain texts, or user-uploaded content, without giving exact file paths or storage locations. Such customised model cards would enable rights-holders to self-identify potential uses of their works, balancing the need for oversight and trade-secret protection.

A partnership between rights holders and the AI developers could also be considered to build technical safeguards. An idea would be a public registry of all known copyrighted content or watermarked models. For instance, policy experts propose 'standardising watermarking and maintaining a registry of watermarked models and detection services' whereby users can easily check any content. By this analogy, publishers would be able to register digital fingerprints (hashes) of their works in a common database. AI labs may then use automatic matchers, for example, perceptual hashing or machine-learning classifiers, to flag protected inputs in the training data or outputs against that registry.

Alternatively, consortia of creators may want to get select access to model Application Programming Interfaces (APIs) with copyright-compliance tests or work on digital rights registries, like those used for music. For instance, US collecting society SoundExchange is formulating a worldwide AI registry that would allow rights holders to opt in or out of having their sound recordings used for training. Such collaborative infrastructure could facilitate identifying copyrighted material used in training or created by an AI system, thereby providing rights holders with a method to monitor and enforce their rights besides merely concerning themselves with data collection.

Conclusion

We have seen that the tension between transparency obligations in the EU AI Act and protection of trade secrets represents a challenge. Finding the right balance requires a balanced approach that satisfy both copyright holders' legitimate interests and developers' innovation concerns. High-level dataset summaries, standardised model cards with copyright sections, and collaborative registries may offer practical compromises that honor the spirit of regulatory accountability without exposing proprietary technologies. As implementation proceeds, regulators must ensure trade secret exemptions remain narrowly defined while industry establishes best practices for meaningful disclosure. Only through a measured approach can the AI ecosystem foster both innovation and trust, creating sustainable paths forward for all stakeholders.

This entry was posted on Wednesday, June 11th, 2025 at 11:18 am and is filed under Artificial Intelligence (AI), European Union
You can follow any responses to this entry through the Comments (RSS) feed. You can leave a response, or trackback from your own site.