



June 25, 2025

If you have any questions regarding the matters discussed in this memorandum, please contact the following attorneys or call your regular Skadden contact.

---

**Deborah J. Kirk**

Partner / London  
44.20.7519.7461  
deborah.kirk@skadden.com

**Jonathan Stephenson**

Associate / London  
44.20.7519.7038  
jonathan.stephenson@skadden.com

---

This memorandum is provided by Skadden, Arps, Slate, Meagher & Flom LLP and its affiliates for educational and informational purposes only and is not intended and should not be construed as legal advice. This memorandum is considered advertising under applicable state laws.

One Manhattan West  
New York, NY 10001  
212.735.3000

22 Bishopsgate  
London EC2N 4BQ  
44.20.7519.7000

## CNIL Clarifies GDPR Basis for AI Training – But It’s Just One Part of the Compliance Picture

### Key Points

- The French CNIL’s recent guidance regarding the application of legitimate interest as a legal basis in AI training is welcome, but several other AI regulatory issues remain unresolved.
- Issues such as copyright, database rights, post-training litigation risk and downstream deployment obligations remain governed by other frameworks, including the EU Copyright Directive, the AI Act and national data protection laws.
- Organisations should apply structured judgement at key moments and should resist viewing GDPR training-stage compliance as a solved issue. A well-documented GDPR position remains a key tool for managing AI compliance at scale.
- Foreign issuers would be allowed to sell their stablecoins in the U.S. if the issuer complies with certain conditions, including responding to seizure orders.

---

The French National Commission on Informatics and Liberty’s (CNIL’s) recent guidance<sup>1</sup> on applying legitimate interest as a legal basis for processing under the General Data Protection Regulation (GDPR) to artificial intelligence (AI) model training offers welcome clarity on a contested issue: the use of personal data scraped from public sources.

Although significant, this is only one layer in a complex stack of regulatory issues surrounding the compliant training of AI models and while the CNIL’s guidance helps to reduce uncertainty around GDPR exposure at the training stage, it does not resolve the copyright, database rights, commercialisation or deployment-phase constraints that continue to shape the legality of training AI systems in practice.

### What the CNIL’s Guidance Clarifies – And What It Leaves Open

The CNIL affirms that training AI models on personal data sourced from public content can be lawful under the GDPR’s legitimate interest basis, provided certain conditions are met. This conditions require a credible balancing of interests, demonstrable safeguards, and clear documentation.

---

<sup>1</sup> The CNIL’s [official guidance](#), published in French on 17 June 2025, reflects its final interpretation of how legitimate interest under Article 6(1)(f) applies to AI training. This analysis summarises the guidance in English for practical application.

# CNIL Clarifies GDPR Basis for AI Training – But It’s Just One Part of the Compliance Picture

Some key points of clarification include:

- **Web scraping may be permissible, provided it respects contextual privacy expectations.** Scraping should not occur where sites actively prohibit it (*e.g.*, via robots.txt), or from platforms aimed at minors. The CNIL also cautions against using legitimate interest for meeting recordings or webinars where individuals appear or speak, especially if the content was not clearly intended for reuse or contains sensitive data. Personal blogs, fora and health-related sites may also carry higher expectations of privacy.
- **Training-scale data use is not inherently unlawful.** The CNIL acknowledges that large datasets may be necessary for effective AI development — and thus may be “necessary” for legitimate interest purposes, provided the principles of proportionality and minimisation are observed.
- **End-user benefit may favour the controller in the legitimate interest assessment (LIA).** The CNIL accepts that improvements in accuracy, reliability or functionality may legitimately weigh in favour of processing, subject to a balanced and well-documented assessment.
- **Regurgitation risk must be addressed, not eliminated.** The CNIL does not expect perfection, but does expect evidence of mitigation — such as prompt filtering, exclusion of high-risk inputs (*e.g.*, usernames, forum posts), and internal testing for memorisation or leakage using probes. If training data or outputs reveal sensitive characteristics, such as political views, ethnicity or health, heightened justification and mitigation are expected.
- **Data subject rights may be respected indirectly.** Where model architectures make individual erasure or objection difficult to implement, the CNIL’s guidance allows for alternatives, such as output filtering to block names, audit trail design or documented suppression logic — provided the rationale is recorded.
- **Documentation must be prepared at the time of training.** The LIA and mitigation planning should be complete and available before AI model training begins, not drafted retroactively in the event of regulatory challenge.
- **DPIAs may still be expected.** The CNIL recommends conducting a data protection impact assessment (DPIA) when model training involves large-scale data scraping, novel content types or special category data, even if legitimate interest is the legal basis.

While the CNIL’s guidance provides welcome clarity on how legitimate interest can support GDPR compliance during AI training, it does not attempt to resolve adjacent legal or strategic questions (nor was it intended to). Issues such as copyright, database rights, post-training litigation risk and downstream deployment obligations remain governed by other frameworks, including the EU Copyright Directive, the AI Act and national

data protection laws. The CNIL also recognises that its position is not harmonised across the EU. Alignment at the European Data Protection Board (EDPB) level remains a live and problematic issue.

## How Other Regulators Compare

While the CNIL’s guidance is the most structured to date, other data protection authorities are operating with varying levels of clarity and emphasis:

- **The UK Informational Commissioner’s Office (ICO)** has acknowledged that existing GDPR rules — including legitimate interest — may be sufficient to justify AI training in some contexts. In its 2023 generative AI consultation, it flagged the possibility of relying on legitimate interest but did not issue detailed implementation guidance on when this would and would not be acceptable. Its current emphasis is broader: on explainability, fairness, children’s data and systemic risk.
- **The Irish Data Protection Commission (DPC) and Italian Garante** have focused primarily on deployment-phase enforcement — particularly failures to conduct DPIAs or provide sufficient transparency around profiling. Several AI deployments in Europe have already been paused or delayed due to unresolved GDPR issues, including data subject rights handling and lawful basis clarification.
- **A consistent, pan-EU approach remains absent.** The CNIL’s leadership position may influence upcoming work at the EDPB level, but for now, companies must navigate multiple expectations depending on where their models are trained or deployed. The CNIL’s role is particularly influential given France’s prominence in AI research and development — including companies such as Mistral and Hugging Face, which face direct oversight from the CNIL as their lead authority under GDPR.

## The Broader Landscape: Legal Uncertainty Beyond GDPR

CNIL’s guidance offers a defensible GDPR position for model training, but it does not resolve other legal restrictions that still limit AI system viability, particularly in commercial settings.

- **Copyright and database law remain binding.** Publicly accessible content may still be protected under copyright or *sui generis* database rights. In the EU, the commercial text and data mining (TDM) exception can be overridden via opt-out mechanisms. In the UK, there is currently no equivalent exception for commercial use (and early-stage plans issued in a consultation have attracted significant criticism from IP holders). So even where GDPR lawful basis is established, the use of the dataset may still be infringing.

# CNIL Clarifies GDPR Basis for AI Training – But It’s Just One Part of the Compliance Picture

- **Contractual terms restrict access and reuse.** Many platforms prohibit scraping or commercial reuse of content via their terms of service. These restrictions are enforceable separately from data protection laws.
- **Downstream deployment introduces new layers.** The AI Act (risk classification, provider-deployer duties), the Digital Services Act (recommender systems, transparency), the UK Online Safety Act (content-based profiling), and sector-specific rules (*e.g.*, health, finance, employment) all carry compliance obligations not resolved by GDPR training-stage alignment.
- **Global regulatory uncertainty is evolving.** The EU AI Act has passed, but implementation is phased through 2025–2027. In the US, efforts at federal AI legislation have slowed amid political deadlock. In the UK, the pro-innovation, regulator-led strategy has not yet produced binding requirements — leaving enforcement to the ICO, the Competition and Markets Authority (CMA) and sectoral bodies.

This regulatory fragmentation is increasingly geopolitical. As French President Emmanuel Macron observed in February 2025, “AI can’t be the Wild West ... there have to be rules.” US Vice President JD Vance said at the same Paris conference, “To restrict its development now would not only unfairly benefit incumbents in the space, it would mean paralysing one of the most promising technologies we have seen in generations.”

The CNIL’s guidance reflects a practical application of what exists, rather than a wait for what’s next. In that sense, it offers both legal clarity and a signal of what politically durable AI oversight might look like: documentation-led, risk-based and interoperable.

## Operational Priorities and Legal Positioning

For legal, privacy and product teams navigating these overlapping regimes, the priorities are not about reinventing governance. They’re about applying structured judgement at key moments. That means:

- **Use the CNIL’s guidance to reinforce existing privacy governance.** It helps teams document and justify legitimate interest for AI training, but it should be integrated into a company’s existing LIA, DPIA and risk assessment workflows rather than treated as a standalone governance model.
- **Training-stage compliance does not enable commercial use.** Even where GDPR allows personal data processing, copyright, database rights, TDM opt-outs and platform terms may still prohibit or restrict model training.
- **Deployment remains a separate compliance layer.** If a company’s model is used for profiling, automated decision-making, or targeting, it will still need to address GDPR, Digital Services Act (DSA) recommender transparency and potentially AI Act obligations.
- **Work cross-functionally and efficiently.** Privacy, legal, product and engineering teams should connect at key model inflection points — *e.g.*, when introducing new data types, deploying user-facing features or enabling automated outputs. The goal should be rapid, practical decisions, not added process.
- **Assign internal accountability.** Companies should ensure there is clear ownership for connecting model training decisions to underlying privacy documentation, particularly where training data changes over time or new models are introduced iteratively.
- **Plan for inconsistency — and document everything.** Until the EDPB adopts a harmonised position, national regulators may expect different standards. Where alignment is not possible, companies should create an internal compliance narrative grounded in this guidance, and should be ready to defend it.

Even with this clarity, organisations should resist viewing GDPR training-stage compliance as a solved issue. Interpretation will vary across member states, and enforcement will likely focus on end-to-end outcomes — particularly where sensitive use cases are involved.

As regulators turn to existing frameworks, a well-documented GDPR position — grounded in the CNIL’s guidance -- remains a key tool for managing AI compliance at scale.