

# AI Insights

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

---

## David A. Simon

Partner / Washington, D.C.  
202.371.7120  
david.simon@skadden.com

## William Ridgway

Partner / Chicago  
312.407.0449  
william.ridgway@skadden.com

## Lisa V. Zivkovic

Associate / New York  
212.735.2887  
lisa.zivkovic@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

## Public Consultation Period Closes for French CNIL Guidance on GDPR-Compliant Development of AI Systems

### Executive Summary

On 16 October 2023, France's Data Protection Authority, the National Commission on Informatics and Liberty (CNIL), issued a set of guidelines for complying with the EU General Data Protection Regulation (GDPR) when researching and developing AI systems (the Guidelines).<sup>1</sup> The public consultation period for the Guidelines closed on 16 November 2023.

The Guidelines confirm the compatibility of the GDPR with AI research and development, and present seven sets of instructions (the "How-To Sheets") with necessary steps that organizations should take to develop AI systems in a GDPR-compatible manner.

Noteworthy takeaways from the How-To Sheets include:

- The purpose limitation principle under the GDPR applies to general-purpose AI systems, meaning the overall purpose of development of an AI system must be specific, explicit and legitimate (even if the purpose in the deployment phase is disparate from the purpose in the development phase).
- Consent, legitimate interest, performance of a contract and public interest can theoretically serve as legal bases for the development of AI systems.
- The principle of data minimization does not prevent the use of large datasets. Reusing datasets (such as publicly available datasets) to train AI systems is permissible, provided that the data was lawfully collected and the purpose for reuse is compatible with the original purpose disclosed at collection.

CNIL will publish the Final Guidelines in early 2024.

### How-To Sheets: Guidance on GDPR Compliance

The How-To Sheets provide practical guidance for data protection officers, legal professionals, AI developers and any other actors involved with AI system development (AI Developers) on seven topics: (i) determining the applicable legal regime governing AI system development; (ii) defining the purpose of processing personal data; (iii) defining the legal qualification of AI system providers (*e.g.*, controller, processor or joint controller);

---

<sup>1</sup> CNIL's statement is part of a wider trend of data protection authorities discussing the interaction between the GDPR and AI development. For example, the [UK Information Commissioner's Office has prepared detailed guidance](#), the [European Parliament commissioned a study](#) on the interaction and [CNIL has generated further guidance](#).

# Public Consultation Period Closes for French CNIL Guidance on GDPR-Compliant Development of AI Systems

(iv) identifying the appropriate legal basis for and ensuring the lawfulness of the data processing; (v) carrying out a data protection impact assessment (DPIA) when necessary; (vi) taking data protection into account in the AI system design choices; and (v) applying data protection principles to the collection and management of learning data.

In so doing, the How-To Sheets attempt to address some open questions about how AI Developers can comply with GDPR principles and requirements in building AI systems.

## Legal Bases for Processing Training Data That Contains Personal Data

Because organizations can only process personal data if they have an adequate legal basis to do so, the use of personal data as training data must also be in accordance with an appropriate legal basis. The How-To Sheets clarify that the legal basis for processing personal data for the AI Development Phase could be:

- **Consent.** Consent can serve as the legal basis for processing personal data as training data when consent is freely given and sought on a case-by-case basis for each specific element of the AI Development Phase where the intended purposes are distinct. However, CNIL notes that situations will arise where consent will not be a suitable ground for processing personal data, including in most employee/employer relationships.
- **Legitimate interests.** (CNIL will produce a specific how-to sheet on this topic at a later date.) Where the infringement of the data subjects' rights to privacy is proportionate to an organization's commercial interests in processing personal data as part of the AI Development Phase, the organization can rely on this ground. However, CNIL notes that where an AI system will be used for profiling, the AI development may fail the proportionality test.
- **Performance of a contract.** Provided the AI Developer can show the necessity of the processing in relation to the object of a contract, the AI Developer can rely on the performance of a contract as a legal basis for processing personal data. However, CNIL specifies that processing users' personal data is not necessary to a social media company's terms of service because the underlying purpose of the terms is to provide access to the social media site, not process personal data for the AI Development Phase.
- **Public interest.** Private and public research institutions using personal data in the AI Development Phase can rely on research in the public interest as a legal basis for the use.

## Purpose Limitation

The purpose limitation principle requires personal data to be processed for a specific, defined purpose. CNIL recognizes that the purpose for processing may not be clearly identifiable in the development phase for certain AI systems (such as general-purpose

AI systems) because the AI Developer may not be able to detail all the future purposes for which the AI system may process the training data (including personal data). In this scenario, the purpose of the developmental processing must identify the following information: (i) the type of system (*e.g.*, a large language model); and (ii) the reasonably foreseeable technically possible functionalities and capabilities of the AI tool. If both have been well defined at the outset of the AI Development Phase, AI Developers may be able to show compliance with the purpose limitation principle.

Conversely, referring only to the type of AI system being designed (*e.g.*, development of a generative AI model) without referring to the technically feasible functionalities and capabilities will not be considered compliant with the GDPR transparency requirement.

## Data Protection Impact Assessments

Where data impact assessments (DPIAs) are required to address the risks of processing personal data for the development of AI systems, the DPIAs must address specific AI risks, such as the risk of producing false content about a real person.

## Data Minimisation and Storage Limitation

The principles of data minimisation and storage limitation restrict the scope of personal data that can be processed and stored to that which is necessary to achieve the purpose of the processing or storage. The Guidelines note that the data minimisation principle does not prevent the use of large datasets containing personal data in the AI Development Phase, as long as the datasets avoid the use of unnecessary personal data and data security safeguards are implemented. Additionally, the Guidelines note that the storage limitation principle does not necessarily restrict long-duration training datasets where such retention is justified with reference to the AI system being developed.

## Reuse of Datasets

The Guidelines note that the reuse of existing datasets in the AI Development Phase, including the reuse of publicly available datasets containing personal data, can be compatible with the GDPR, provided that the collection of the personal data was compliant with the GDPR and the purpose of reuse in the AI Development Phase is compatible with the purpose of the initial collection.

## Conclusion and Steps for AI Developers To Consider

CNIL's How-To Sheets provide practical guidance for AI Developers when using personal data to train AI systems in ways that are compatible with the GDPR. Additional guidance from CNIL is due in the coming months. The Guidelines show a regulatory willingness to engage actively with AI Developers to allow safe and GDPR-compliant innovation at the outset of AI development projects. AI Developers should review their use of personal data for alignment with the practical examples set out in the How-To Sheets.