



Title: DESIGNING AI SYSTEMS IN COMPLIANCE WITH THE GDPR

Brief Overview:

For AI systems containing personal data, developers must ensure that their model complies with the GDPR to prevent the risks faced by data subjects. Some professionals having reported their difficulties to ensure compliance of their AI systems, the CNIL has recently published a set of Guidelines to help developers check compliance with the GDPR.

Designers and developers of artificial intelligence systems often report to the CNIL that applying the GDPR raises significant challenges for them, particularly in terms of training models, which requires a substantial volume of data.

In this context, the AI Act and the GDPR have to be considered as complementary sets of rules. Recitals of the AI Act indicate that this Regulation is to be implemented without prejudice to the GDPR. In addition, the GDPR sets out the principle of technological neutrality which ensures its application to any technology, such as an AI system, that processes personal data.

Thus, AI systems must be created and developed in a manner that ensures the protection of individuals, in accordance with the principles of privacy by design and by default set out in Article 25 of the GDPR.

On July 22, 2025, the French data protection authority (CNIL) published 27 pages of Guidelines to help professionals to ensure GDPR compliance of their AI systems. These guidelines focus solely on the development phase during which the system is designed, the database is created, and the learning process is implemented. They provide professionals with a list of 12 items to check and carefully document, in the following order:

1. **Define the purpose(s) of the processing** for each category of data collected
2. **Identify the responsibilities** of each party involved in the creation of the AI system, prior to its development, as data controller or data processor. This allocation of responsibilities is determined on a case-by-case basis, depending in particular on who selects the AI training data, and must be specified in their contract.
3. **Determine the legal basis that authorizes the processing** : Developing an AI system requires a substantial database. However, obtaining consent from every person whose data is collected can be difficult in practice, if not impossible when data is provided by a third party or from an unrestricted database. According to the Guidelines, one solution would be to rely on the legal basis of legitimate interest, especially when web scraping is used, provided that the three conditions established by the CNIL (*see Guidelines “Legitimate interest: how to base processing on this legal basis?” published on December 2, 2019*) are met and suitable guarantees are implemented such as a discretionary right to object prior to collection, anonymizing or pseudonymizing the data, and excluding websites containing particularly sensitive data.
4. **Verify whether the collected data can be further processed** : this involves ascertaining the origin of the initial collection (directly from the data subject by the data controller or from third parties or unrestricted databases) which defines the rights and limits applicable to the initial collection and therefore the legality of a further processing.
5. **Respect the principle of data minimization** when creating a large database which requires defining in advance the categories of data needed to train the AI according to its intended use, then collecting only these categories from sites that do not refuse automatic collection, and finally deleting unnecessary data after verifying all the data collected.



6. **Set out a data retention period** for the data used to create the AI system's training database. This period will depend on the category of data, the legal basis used, and the purpose of the AI system (general or specific. When this data is no longer useful for training the AI system, it must be deleted unless safeguards are implemented (restricted access to authorized persons only, partitioned storage, etc.) to use it for product maintenance or improvement purposes.
7. **Inform the data subjects**, within a reasonable time between the data collection and the training, notably about the processing of their personal data, the retention period, their rights and how to exercise them, but also the risks associated with processing personal data through an AI system and the measures taken to prevent them. According to the GDPR, this information must be communicated individually to data subjects, unless it has already been provided to them or doing so would require a disproportionate effort. The latter scenario may arise in the case of web scraping of pseudonymized data, given the effort required to identify and contact the data subjects.
8. **Ensure the exercise of rights of the data subject** over the training database if the data subject can be identified even though its data had been pseudonymized or anonymized
9. **Implement security measures**, such as control of access to data, encryption of backups and communications, an AI shutdown system, etc., to guarantee the confidentiality and the integrity of the training data.
10. **Analyze the status of the AI model**, whose training database contains personal data, to determine whether the AI model is subject to the GDPR or if the model can be considered anonymous. Status analysis consists of attempting to re-identify data extracted from the database using, for instance, attack tests or other means that could be used against the model.
11. **Labelling each data**, which consists of assigning a description to each data so that the AI system can easily recognize and use the data to respond to the requests submitted.. The labelling must be accurate, objective, relevant and limited to the very data necessary for the purpose of AI training, in accordance with the principles of minimization and accuracy.
12. **Carry out a data protection impact assessment** (Article 35 of the GDPR) in order to map and assess the risks associated with the processing of personal data by the AI system and thus establish solutions and measures to protect the data subjects.

Compliance with the current version of the AI Act is compulsory even though a simplification of the rules had been presented by the European Commission on November 19, 2025 though its new digital omnibus. However, it will only come into force once, and if, it has been adopted by both the European Parliament and the Council.



Authors:

- Frédéric Lecomte, Partner, Flecomte@bersay.com
- Bradley Joslove, Partner, Bjoslove@bersay.com
- Léa Paravano, Associate, Lparavano@bersay.com
- Guka Otkhmezuri, Associate, Gotkhmezuri@bersay.com