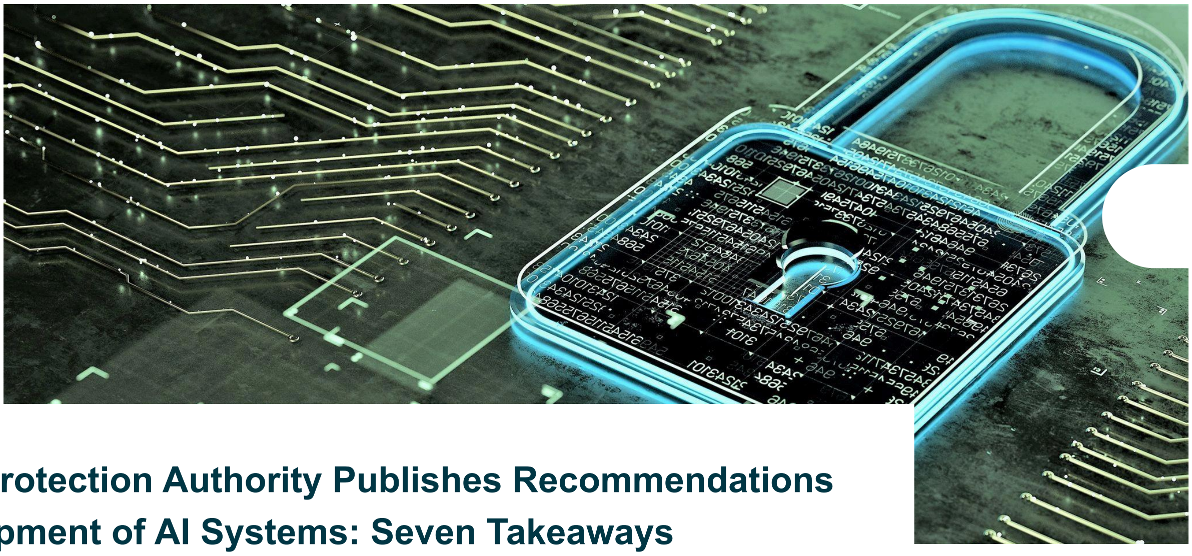

French Data Protection Authority Publishes Recommendations on the Development of AI Systems: Seven Takeaways

ALERTS

April 23, 2024

On April 8, 2024, the French Data Protection Authority (**CNIL**) published **recommendations on the development phase of artificial intelligence (AI) systems**[1] (**Recommendations**). They are the first set of recommendations designed to guide the various players in the AI ecosystem on how to apply the General Data Protection Regulation (GDPR) to the development of AI systems. The Recommendations are relevant to providers and users of AI systems who process personal data as part of the development of such systems, including fine-tuning already-existing AI systems.

Below are seven takeaways from the Recommendations:

1. **Your role under the AI Act does not exempt you from your GDPR responsibilities.** Roles and responsibilities under the AI Act and GDPR differ and apply concurrently. While the AI Act classifies organizations based on their role as providers, importers, distributors, and users of AI systems, GDPR classifies them as either controllers (i.e., they determine the purposes and means of the processing) or processors (i.e., they process data on behalf of a data controller). According to the CNIL:
   - AI system providers should be considered **data controllers** if they initiate the development of an AI system and build their training databases from the data they have selected. Organizations can also be considered **joint controllers** if they are involved in the building of a training database together with a provider.
   - In other cases, providers of an AI system should be treated as **data processors** (e.g., if they develop an AI system on behalf of one of their customers).
2. **Know why you are developing an AI system.** The CNIL notes that providers should be clear about why they are developing their AI system. Any processing of personal data used to develop an AI system should have a specific, explicit, and legitimate purpose (i.e., a clearly defined objective). The CNIL distinguishes between:
   - **AI systems which have a specific purpose from the start.** In this case, both the development phase and the deployment and use phase have the same purpose. An example would be an AI system used to sort job candidates' resumes.
   - **General-purpose AI systems.** These systems can be used in diverse contexts, making defining a sufficiently specific purpose harder. The CNIL considers a purpose sufficiently specific if it refers to both the type of system being developed (e.g., a large-scale language model or a computer vision system) and the functionalities and capabilities that are technically foreseeable.
   - **AI systems that are developed for scientific research purposes.** In these cases, the purpose can be less detailed, given the difficulties of defining it at the beginning of the research activities. As the research project progresses, the purpose of the processing can be clarified further.

CONTRIBUTORS

Yann Padova

Cédric Burton

Tom Evans

Marie Catherine Ducharme

Mia Gal

3. **Rely on a valid legal basis.** Any development of an AI system involving the processing of personal data should have a legal basis. The legal bases that can be relied on when putting together a database to train an AI system are the following: consent, legitimate interest, public interest, contract, and legal obligation. The CNIL indicates that it might be complicated to obtain consent in certain cases (e.g., if a controller is reusing an open database) and considers that legitimate interest can be a valid legal basis provided that:
   - The interest pursued by the data controller is "legitimate." In most cases, if the purpose of an AI system is legitimate, then the processing of personal data for the purpose of the AI system's training will also be legitimate.
   - The processing is necessary to achieve the interest. For example, this means that the processing cannot be carried out with anonymous or synthetic (artificially generated) data.
   - The interests and rights of the data *subjects "are not disproportionately prejudiced."* This means that sufficient safeguards have been implemented to limit the impacts of the processing on individuals (e.g., carefully reviewing the categories of data processed and removing any data not strictly required for the training of the AI system).

4. **If you are reusing personal data, ensure you have the right to do so.** If you are a data controller and are reusing personal data obtained in a different context or for a different purpose, the CNIL recommends carrying out additional checks to ensure that this new processing is compliant with the GDPR. The CNIL addresses the following situations:
   - **You are a provider that reuses personal data you previously collected for another purpose.** You should check whether this onward processing is compatible with the initial purpose for which you collected the data.
   - **You are a provider that reuses publicly accessible data.** While it is primarily the data controller who made the data public who should ensure that it did so lawfully, the data controller reusing such data must ensure that it is not relying on a clearly unlawful database (e.g., coming from a data breach). The CNIL advises reusers to ensure the database cites its source, does not encompass criminal offense records, and raises no clear doubts about its legality.
   - **You are a provider that reuses data obtained from third parties.** The third-party sharing personal data should ensure that such sharing is legal, and the reuser should perform checks to ensure that the database they are reusing is not clearly unlawful.

5. **Conduct a Data Protection Impact Assessment (DPIA) when required: high risk under the AI Act versus high risk under the GDPR.** In certain cases, conducting a GDPR DPIA when deploying an AI system will be necessary. This obligation is separate from the compliance documentation required under the AI Act. The CNIL provides some pointers on when to conduct a GDPR DPIA:
   - **If an AI system is classified as "high risk" under the AI Act, a GDPR DPIA will be required**. The provider of an AI system may reuse the documentation prepared for the purpose of complying with the AI Act to prepare the GDPR DPIA.
   - **If an AI system is not classified as "high risk" under the AI Act, a DPIA will be required when the processing of personal data results in a "high risk" for the rights and freedoms of individuals**. The existence of a "high risk" should be considered on a case-by-case basis. Elements to consider include whether the processing involves special categories of data ("sensitive data"). Interestingly, the CNIL provides that not all AI systems will constitute "innovative use" (for example, technology that has already been tried and tested for many years in real-world conditions), while this is a criterion often used to determine whether to conduct a DPIA.

6. **Data minimization and proportionality are key.** When considering the design choices to be made for an AI system, you should comply with data protection principles such as data minimization and proportionality. According to the CNIL, the principle of data minimization does not prevent AI systems from training based on very large volumes of data. However, it requires identifying the personal data needed to develop the system in order to avoid using unnecessary personal data, whether in the training data, the associated metadata, or the annotations and features. Starting with a pilot project and/or referring to an ethical committee is also good practice.
7. **Plan the data lifecycle: select, archive, and delete.** As for any processing of personal data, retention periods and access rules should be set. Different access and retention periods should apply to, on the one hand, the building of the database and the training of an AI system and, on the other hand, the maintenance (including for audit and bias measurement) and improvement of the AI system. Companies should also consider measures like pseudonymization and data segregation.

Next Steps

Companies should consider the above when training AI powered solutions involving the processing of personal data that they maintain or procure. More guidance is also coming: the CNIL indicated that it will issue a second set of recommendations on the deployment phase later.

For more information, please contact Cédric Burton, Yann Padova, Laura De Boel, or another member of the firm's privacy and cybersecurity practice.

Mia Gal, Marie Catherine Ducharme, and Tom Evans contributed to the preparation of this Wilson Sonsini Alert.

---

[1] https://www.cnil.fr/fr/ia-la-cnil-publie-ses-premieres-recommandations-sur-le-developpement-des-systemes-dintelligence