BCLP. Client Intelligent

Insights

ANALYSING THE CNIL'S LATEST RECOMMENDATIONS FOR AI SYSTEMS

Jun 27, 2024

SUMMARY

Following the very recent adoption of the EU Regulation on AI (the AI Regulation) the CNIL (the French data regulator) has issued the second in its series of recommendations for the development of privacy-friendly AI models.

This forms part of the CNIL's work to make players in the AI ecosystem more accountable, following its issue of its first recommendations on 8 April 2024. Although these new recommendations are open for public consultation until 1 September 2024, they provide for a useful guidance for AI systems developers, as well as an early indication of the CNIL's regulatory direction of travel. And in developing this guidance, the CNIL is also establishing itself as the most advanced supervisory authority in Europe in terms of the adoption of a regulatory regime designed to promote the development of AI whilst respecting personal data rights.

IS LEGITIMATE INTEREST AN APPROPRIATE LEGAL BASIS FOR PROCESSING PERSONAL DATA FOR AI DEVELOPMENT?

The CNIL indicates that can be, but it will constitute a valid legal basis only if:

- The interest pursued is genuinely legitimate;
- The processing is justified by the condition of necessity;
- Processing does not disproportionately prejudice the rights and interests of the persons whose data is processed.

THE INTEREST PURSUED IS LEGITIMATE

Data controllers must precisely identify the interest being pursued and inform data subjects. The CNIL has suggested that the following interests may constitute legitimate interests in the context of the development of AI systems:

- Carrying out scientific research,
- Facilitating public access to certain information,
- Developing new systems and functions for users of a service,
- Offering a conversational agent service to assist users,
- Improving a product or service to enhance its performance,
- Developing an AI system to detect fraudulent content or behaviour.

This will not be the case however, if the interest is not linked to the organisation's mission or if the system cannot be deployed legally (for example, an AI system that automatically profiles minors in order to send them targeted advertising or deployment of an AI systems that is prohibited by other regulations, such as the AI Act).

THE PROCESSING IS NECESSARY

The data controller must ensure that the development of the AI system is in fact necessary to achieve the objective pursued.

AVOIDING DISPROPORTIONATE PREJUDICE TO THE DATA SUBJECT

This proportionality test should be conducted at each of the following stages:

- **Data collection**: for example, *web scraping* can pose risks to privacy and freedom of expression,
- **Model training and data storage**: to assess possible risks posed by the loss of data confidentiality resulting in a potential breach of privacy, the lack of transparency and opacity of processing, and the difficulty of guaranteeing the exercise of rights,
- **Use of the AI system:** the data controller must implement sufficient safeguards to minimize the risks of discriminatory bias and illicit re-use of the data.

DUTY TO INFORM

Whilst in principle, information on the retention period and the exercise of rights does not have to be provided as a matter of course, in practice, it will almost always be required when training data is created and used. When data is collected indirectly, AI system providers must also inform data subjects of the categories of personal data and their source(s) – whether publicly accessible or not:

- If the dataset has been reused, the new controller should provide a means of contacting the controller from whom the data was retrieved, and a hypertext link to the website of the initial controller and a summary of the conditions of collection.
- In the case of web scraping, the CNIL recommends providing – at least – the categories of source websites concerned, and if possible the domain names and URLs of the web pages concerned. This obligation must be seen in tandem with Article 53 of the AI Act, which requires providers of general-purpose AI models to make available to the public a sufficiently detailed summary of the content used to train the model.

The CNIL also considers it good practice for the supplier to specify the following:

- The nature of the risk associated with the reconstruction of data from the model, such as the risk of data regurgitation in the case of generative AI;

The measures taken to limit these risks, and the availability of recourse mechanisms (such as the ability to notify the organisation of an instance of regurgitation).

RIGHTS OF DATA SUBJECTS

Data subjects must be able to exercise their GDPR rights with regard to training databases and AI models. In practice, the response to such requests will vary depending on whether these relate to the training data or the model itself.

TRAINING DATA

Difficulties in identifying the data subject

Where the controller does not or no longer needs to identify the data subject and can demonstrate that it is unable to do so, it does not have to retain or collect additional information for the sole purpose of enabling a data subject to exercise his/her rights. The controller should inform the data subject that it is not able to identify the data subject, if possible. Data controllers should however anticipate these identification difficulties and inform the data subjects of the identifying data needed to re-identify them (as part of its duty to inform data subjects).

Right to receive a copy of learning data

The right of access implies a right for the data subject to obtain extracts from the database where this is indispensable for the effective exercise of rights. The CNIL therefore recommends that annotations and associated metadata also be provided in an easily understandable format.

Database modification

A data subject may be able to object to processing based on legitimate interests for reasons relating to the particular situation of the data subject. Where data can be scraped from websites, it will be good practice to set up a list, managed by the data controller, enabling data subjects to object to the collection of their data on certain websites or platforms, including prior to collection.

AI MODELS SUBJECT TO THE GDPR

Generative AI models

The outputs of GenAI models may constitute personal data. However, the supplier of the GenAI system will not be responsible for the processing of personal data contained in the outputs which do not result from storage by the model but from statistical inference. This processing will be the responsibility of the system user only. Where GDPR applies to the model, the data controller may still be able to demonstrate that it is unable to identify individuals within its model. More often than not, the data controller will be able to demonstrate that the state of the art does not allow personal data to be identified on the basis of the model's parameters.

Information on the processing by the model and the right to obtain a copy of the data

Where the data controller has been able to identify the data subject and verify that data storage has taken place, this must be confirmed this to the data subject. Where it has not been possible to verify that data has been stored, but where the data controller has not been able to rule out the possibility of such storage, the CNIL recommends that data subjects be informed that it is not impossible for their data to be comprised in training data stored by the model.

Exercising rights to rectification, opposition or deletion of data in the model

When the data controller still has the training database, the model can be re-trained with the relevant data subject’s personal data removed from the database.

As current technical solutions enabling identification and removal of personal data are not of a satisfactory standard for AI models subject to GDPR, the CNIL suggests that data controllers should observe a reasonable period of time between the time the learning database is created and the time when the model trains itself or the dataset is disseminated, to allow data subjects to exercise their rights upstream.

ANNOTATING DATA

Data annotation makes it possible to assign a description to each piece of data, which is used as a *ground truth* to enable the model to process, classify distinguish data on the basis of this information. Where annotation involves personal data, this must be carried out in compliance with the GDPR.

MINIMISATION

The minimisation principle will not be met if the annotations contain information that is not relevant to the intended functionality. Information will only be considered relevant if its link with the model's performance is proven or sufficiently plausible. When the annotation activity is entrusted to a third party:

- if the data set is configured specifically for the client's needs, the supplier of the data set will be a sub-processor. The supplier must therefore ensure that the dataset contains only relevant annotations;
- if the supplier makes available training data sets that have already been created, it must ensure it has designed its product so as to ensure compliance with the minimisation principle.

ACCURACY

For the purposes of compliance with the accuracy principle, annotations must only contain accurate information about the person to whom the data relates. The developer must therefore take appropriate measures to ensure that the principle of accuracy is respected.

QUALITY OF ANNOTATION

The CNIL recommends that AI model providers:

- define a continuous verification procedure by defining an annotation protocol and a continuous verification procedure (e.g. sample analysis, audits); and
- involve an ethics representative or committee, upstream and during the annotation phase.

INFORMING DATA SUBJECTS

Data subjects must be informed of the annotation process (including its purpose, the person responsible and relevant security measures). Again as a matter of good practice, the data subjects may also be informed of the results of the annotation - particularly when the annotation is likely to

have consequences for the data subjects. Data subjects can also exercise their access rights in relation to annotations.

ANNOTATION BASED ON SENSITIVE DATA

Annotation may sometimes reveal sensitive data without the source data itself being sensitive data. It will be up to the data controller to identify one of the legal bases in the GDPR so as to carry out the processing legally. In any event, the CNIL recommends that such data should only be processed in very exceptional circumstances, and that preference should be given to other categories of data. Data controllers should only annotate according to objective and factual criteria and must increase the security of annotated data.

SECURITY

In AI system development, developers must combine a "classic" security analysis focusing in particular on the security of the IT and software environment with an analysis of the risks specific to AI systems and large-scale training databases. The development of an AI model generally requires a data protection impact assessment. This analysis should guarantee data confidentiality (by considering risks of data storage, reconstruction or inference of membership) and system performance and integrity, as well as ensuring the overall security of the information system. To assess the level of risk presented by the AI system, AI model providers must take into account, in particular, the nature of the data, control over the data and tools used, how the system is accessed, the content of the system's outputs and the intended context of use for the AI system.

SECURITY MEASURES

The CNIL has drawn up a list of recommended security measures, dependent on the phase of depending on the training, development or operating phase of the system. This focuses on checking reliability and quality of training data, considering encryption and data access controls, as well as use of fictitious or synthetic data and whether it is possible to anonymise or pseudonymize data. Developers of training data sets may also want to consider how to prevent loss of control through organisational measures (e.g. by watermarking data).

When developing the system, the CNIL recommends following good security practice in the field (by using libraries, tools or pre-trained models that have already been verified) and using a controlled, reproducible and easily deployable development environment, implementing a continuous development and integration procedure.

When the system is operational, the developer of the AI should:

- inform the user of the limitations of the system in the intended contexts of use;,
- provide information enabling the user to interpret the results;
- provide a means of stopping use of the system; and
- check system outputs.

The BCLP Global Data Protection team is available to guide you through the process of developing your AI systems or models, whether in the EU or elsewhere. Please contact Pierre-Emmanuel Frogé or your usual BCLP Data contact.

RELATED PRACTICE AREAS

- Data Privacy & Security

MEET THE TEAM

Pierre-Emmanuel Froge

Paris

pierreemmanuel.froge@bclplaw.com
+33 (0) 1 44 17 76 21