Public consultation – AI how-to sheets on building datasets for the design of AI systems

Summary of contributions

February 2024

On 11 October 2023, the CNIL launched a public consultation on building training datasets for artificial intelligence systems.

The contributions fed into the work on **this first batch of eight how-to sheets** with a view to their final publication on the website.

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

Summary in figures

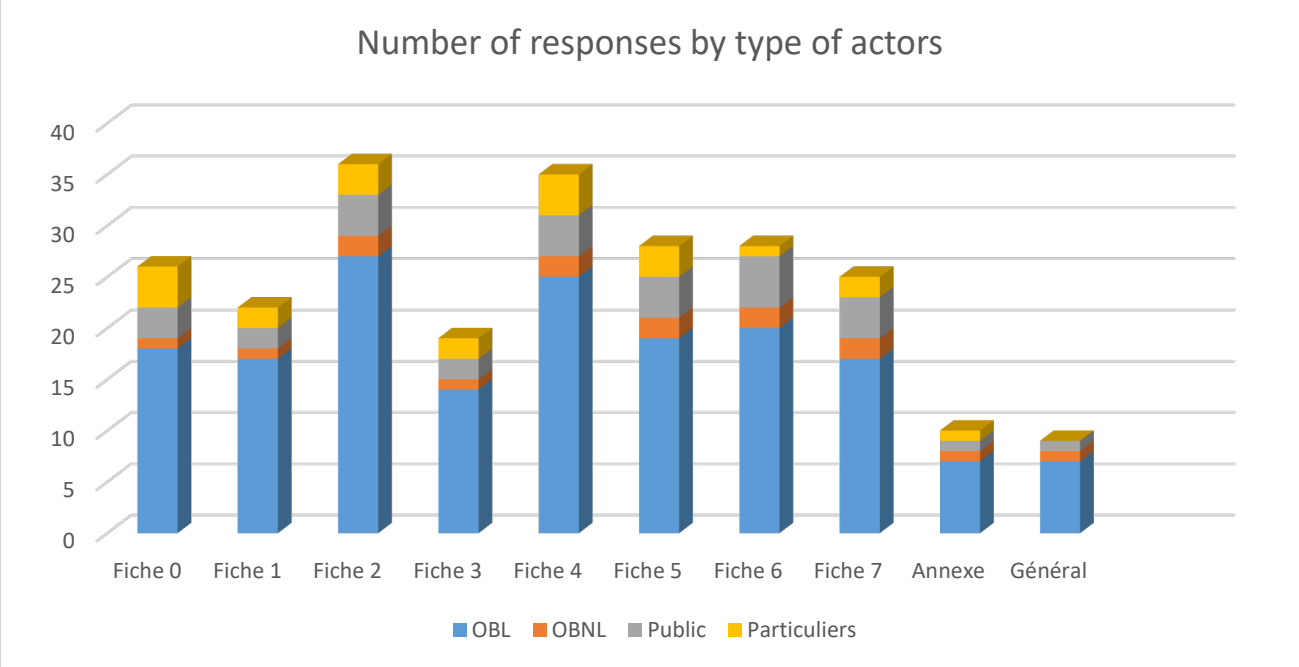
On 11 October 2023, the CNIL published a first set of eight how-to sheets on the creation of training datasets for AI systems, in order to support actors in their efforts to comply with the legislation on the protection of personal data and to answer their main questions.

Following the public consultation, on 15 December 2023, **the CNIL received 43 contributions,** from contributors representing different sectors:

- for-profit organisations from different sectors (AI, finance, health, aeronautics, operators of online platforms, online advertising, video games, audiovisual, etc.):
  - 10 organisations representing professionals;
  - 18 private companies;
  - 1 consulting firm

- non-profit organisations:
  - 1 representative association of civil society;
  - 2 research institutes and 1 grouping of public research institutes;
  - 1 independent think tank;
  - 3 trade unions of employees.

- 4 individuals;
- 3 public institutions.



These contributions enabled the CNIL to:

- **develop its how-to sheets by providing additional clarifications and consolidate its analyses in the light of the comments made in the contributions;**
- **address, in the summary below, the concerns most frequently shared by contributors.**

General contributions on all sheets

The legal value of the how-to sheets

Summary of contributions

Several contributors questioned the legal value of the how-to sheets and the binding or non-binding nature of the recommendations made therein.

CNIL' response

The introductory sheet has been completed to underline that these sheets **recall the obligations imposed by the regulations** (e.g. "*the controller must* (...)") **and make recommendations to comply with them** (e.g. "*may allow*"). However, these recommendations are not binding: controllers may deviate from them, provided that they can justify their choices and under their responsibility.

Some recommendations are also made as good practices (e.g. "*recommended as good practices")*. The monitoring of these good practices goes beyond what the regulations require.

It also makes a clearer distinction in the how-to sheets between what is a reminder of the applicable obligations, compliance recommendations or simply good practices.

On the link with the European AI Act

Summary of contributions

Many contributors consider that it would be useful for the how-to sheets to refer more explicitly to the EU AI Act in order to clarify the articulation.

The problems of articulation identified by the stakeholders concern in particular the following subjects: the definition of AI systems, the qualification of actors, the definition of risks, documentation obligations and the processing of sensitive data for the detection and correction of biases.

CNIL' response

Pending its entry into force, the CNIL recalled **in the introductory sheet that the GDPR already applies independently of the AI Act.** Besides both texts do not share the same scope. It clarifies, where relevant, **certain points of articulation** (in particular on the qualification of actors and the need for a DPIA for systems identified as high-risk in the AI Act).

On the development and deployment phases

Summary of contributions

Some contributions question the choice of division proposed between the development and deployment phases.

CNIL' response

This division allows a chronological separation between the different phases of processing of personal data, which correspond to separate processing (possibly implemented by different bodies) and thus different sets of data subjects.

The CNIL is aware that an important part of the developments currently carried out in AI may consist in implementing them directly or adapting them to specific use cases (by transfer learning or fine-tuning for example). These practices complicate the division into two phases proposed by the CNIL. However, the CNIL cannot exclude the case where a supplier designs a model entirely, whether in order to develop a system directly used in the deployment phase or a general (or foundation) model that will subsequently be adapted by the supplier or a third party. framing this step is crucial because it involves the largest amount of data (since it consists of initializing the model) thus presenting risks for the data subjects.

Introductory how-to sheet - The scope of the AI how-to sheets

The choice of the definition of AI

Summary of contributions

The scope of the systems concerned is intended to encompass all tools using personal data – thus subject to the GDPR – and using AI-like techniques. Contributors asked if the definition chosen corresponds to that of the AI Act.

CNIL' response

The CNIL proposes to clarify its scope by explicitly using the definition now being adopted in the AI Act.

The inclusion of the fine-tuning in the scope

Summary of contributions

Some contributors asked for the fine-tuning to be included in the scope of the sheets, or even for the processing phases (development/deployment) to be broken down according to the use of pre-existing models.

CNIL' response

The CNIL has included fine-tuning or transfer learning within its scope: they can be assimilated to processing in the development phase.

While fine tuning an existing model is a common use case in AI, **it does not call into question the splitting into two phases of development and deployment of an AI system.** Fine tuning, like the development of a model from scratch, requires design choices, the collection of a dataset and a learning phase. An organization performing fine-tuning will have a system provider role for the model it has designed and will be bound by the same obligations.

A difference should be noted with regard to the pre-existing model used as a basis for fine tuning. This model will come from an earlier development phase after which it will have been transmitted or made available (e.g. open source). In the absence of specificities at the development stage, questions relating to the transmission or availability of the pre-trained model will only be addressed at a later stage.

The risks of the presence of personal data

Summary of contributions

Some contributors asked for clarification on how to identify the risks of personal data being present in a dataset.

CNIL' response

The presence of personal data constitutes a risk for data subjects which cannot be assimilated to the concept of risk as defined by the AI Act. Although there are many indications for assessing the presence of personal data, and it is not possible to list them exhaustively, clarifications have been made on the distinction between datasets possibly containing personal data, and those most likely containing personal data. The CNIL thus acknowledges that, since the volumes mobilized for the development of AI systems are potentially very large and the data used are complex (unstructured text, audio, video, etc.), it is in many cases very difficult to be certain that no personal data are included in a dataset.

The distinction between "AI model" and "AI system"

Summary of contributions

Several contributions stressed the need to distinguish between the concepts of "model" and "system".

CNIL' response

That distinction, which overlaps with the distinction between "foundation models" and "general purpose AI systems" made, inter alia, during the RIA negotiations, was repeated where relevant. Thus, the CNIL considers the model as the product of the training carried out on the basis of the training dataset and the system as the software integration of the model which can then be deployed or fine-tuned.

However, the CNIL considers that this distinction is not structuring with regard to the compliance with the GDPR of the processing of personal data in the development phase.

How-to sheet 1 – Determining the applicable legal regime

Summary of contributions

Many contributions show a **lack of understanding of the purpose and scope of the sheet,** particularly in the light of the following elements:

- the absence of a definition of "legal regime";
- the absence of any indication as to the applicability of the GDPR or the Law enforcement Directive;
- confusion between the different cases presented.

CNIL' response

The purpose of this sheet is to assist the controller in determining the applicable data protection regulations (GDPR, Law enforcement Directive or processing operations concerning national defense or state security) when developing AI systems.

In response to those observations, **the CNIL clarified the concept of "legal regime" and included additional illustrations with regard to general purpose AI systems.**

How-to sheet 2 – Defining a purpose

The criteria for defining the purpose for general purpose AI systems

Summary of contributions

Several contributors questioned the relevance of purpose definition criteria for general purpose AI systems:

- some contributors consider that these criteria are too flexible and that they do not effectively meet the criterion of precision of purpose, which could undermine compliance with the principles deriving from them (in particular data minimisation and purpose limitation);
- another part of the contributors consider that these criteria are too prescriptive, especially given the impossibility of predicting the uses of AI systems from the development phase.

CNIL' response

The CNIL has clearly identified the complexities involved in defining a sufficiently precise purpose.

In response, it recalls, in the final version of this how-to sheet, criteria to take into account the difficulties for the controller to define, at the stage of the development of an AI system, all its future applications, while ensuring that the principle of purpose is respected.

In order to ensure legal certainty for stakeholders, **the CNIL has also clarified the distinction between recommendations that may or may not fall within the scope of good practice.**

Re-use of scientific research data for other purposes

Summary of contributions

Several contributors called for clarification of the concept of scientific research and **the conditions under which it is possible to reuse data from scientific research for other purposes, including commercial ones.**

CNIL' response

The re-use of anonymized data or of a model that has not stored personal data, even outside a research purpose, does not pose any difficulties.

On the other hand, the CNIL recalls, in the final version of how-to sheet 4,the principle that **the re-use for non-research purposes of personal data initially processed for research purposes is lawful only for purposes deemed compatible.** This new processing will have to comply with all the principles laid down by the GDPR (information of individuals, respect for rights, identification of a new exception for the processing of sensitive data where applicable, etc.).

How-to sheet 3 – Determining the legal qualification of stakeholders

On the link with the AI Act

Summary of contributions

Several contributions called for a more explicit clarification of roles and qualifications within the meaning of the GDPR in conjunction with the European AI Act.

CNIL' response

Although the AI Act has just been adopted, developments have been added to clarify the qualifications that an "AI system provider" could assume within the meaning of the GDPR. In addition, an example now illustrates the particular case of the model fine-tuning that would have memorized personal data.

The **dissemination, storage or maintenance of such a model is regarded as the processing of personal data.** This has consequences in terms of responsibilities, in particular for the provider of such AI models.

The value of the criteria and examples given

Summary of contributions

Some contributions felt that the criteria and examples of qualification were too prescriptive.

CNIL' response

The **legal qualification of providers of AI systems must be made on a case-by-case basis.** This how-to sheet is not intended to create new criteria, but rather to shed light on the clues to carry out this analysis, already mentioned in the European Data Protection Board's Guidelines 07/2020 on the concepts of controller and processor.

How-to sheet 4 – Ensuring that the processing is lawful

The legal basis of the legal obligation

Summary of contributions

Several contributions question the possibility of mobilizing the legal basis of the legal obligation (Article 6.1.c GDPR) for the development and use of AI systems for certain purposes. Several examples are given (fight against money laundering and financing of terrorism or content moderation).

CNIL' response

The CNIL has completed these sheet to specify the limits of the legal obligation **as a relevant legal basis for system development. However, this** does not preclude the use of an already developed AI system to fulfil a legal obligation when the conditions for mobilizing the legal basis are met.

The incidental processing of sensitive data

Summary of contributions

Several contributors question the **incidental presence of sensitive data in training datasets.** They highlight the difficulty or even the impossibility, in some cases, of ensuring the absence of sensitive data, in particular when the dataset consists of online data collection.

CNIL' response

The processing of sensitive data is, in principle, possible only on the basis of one of the exceptions exhaustively listed in Article 9(1) of the GDPR.

Please note: a clarification is added on the conditions identified by recent case-law (ECJ, 4 July 2023, case C-252/21) for mobilizing the exception relating to the collection of data "manifestly made public".

The CNIL has specified the rules applicable in the event of incidental collection of sensitive data when using web scraping tools for the constitution of AI datasets:

- The controller is obliged to implement all measures to automatically exclude the collection of non-relevant sensitive data. In particular, it must apply filters preventing the collection of certain categories of data and/or refrain from collecting data on certain websites with inherently sensitive data.
- If, despite the measures taken, the organization processes incidentally and residually sensitive data that it had not sought to collect, this shall not be considered illegal. On the other hand, if the organization becomes aware that it is processing sensitive data, it is required to carry out, as far as possible, their immediate and automated deletion.

Additional checks in the event of re-use of freely accessible datasets

Summary of contributions

Some contributors consider that the recommendations concerning the checks to be carried out to ensure that a dataset has not been the subject of a court decision prohibiting its re-use are too restrictive. Several contributors also asked to clarify the issue of proving the absence of "flagrant doubts" on the lawfulness of the dataset and to clarify the detailed procedures for verifying the legality of re-used datasets.

CNIL' response

Manifest wrongfulness must be assessed on a case-by-case basis. As a result, the CNIL considers that it will be for the controller to carry out the necessary checks as appropriate.

How-to sheet 5 – Carry out an impact assessment if necessary

On the link with the proposal for a European AI Regulation

Summary of contributions

Several contributors pointed out that the risk-based approach of the AI Act was not the same as that leading to the realization of a DPIA. In particular, it was requested to clarify whether a DPIA was mandatory for high-risk systems whose development requires the processing of personal data.

CNIL' response

In line with the position taken by the CNIL and its counterparts in the joint opinion published with the European data protection supervisor, it is clarified that **a DPIA would be mandatory for systems classified as high risk by the AI Act**.

Further clarifications on the need for a DPIA for foundation models and general purpose AI systems were also provided.

Finally, in response to questions from several contributors, clarification was provided concerning **the relationship between the documentation requirements of the AI Act and the implementation of a DPIA,** since many elements may be common to both productions.

The risk assessment of third-party models

Summary of contributions

Some submissions raised that it may be difficult to carry out a DPIA for designers of systems based on pre-trained third-party models (which they adapt to their needs through fine-tuning for example).

CNIL' response

The CNIL considers that the recommendation, made to model designers, to carry out a DPIA in order to transmit it to re-users, and in particular to those wishing to integrate them into their own development phase (in particular for fine-tuning) is **a sufficient measure to enable users of pre-trained models to carry out their own DPIA.**

Furthermore, the CNIL recommendation inviting model designers to make their DPIAs public **must provide guarantees on the transmission of the information necessary for users to carry out their own DPIAs in the case of models published in open source.** To be in line with the expectations of the CNIL, these DPIAs must be sufficiently comprehensive to allow users to assess the risks associated with the use of the model in their processing.

The criteria requiring the performance of a DPIA

Summary of contributions

Several contributors argued that these criteria, including "innovative uses" and "large-scale processing", could be difficult to assess without an exact threshold. They invite the CNIL to provide these clarifications.

CNIL' response

The CNIL considers that these thresholds cannot be identified in general for all the processing operations concerned. It falls within **the assessment of the controllers,** who will have to take into account the specific context of their processing.

How-to sheet 6 – Taking data protection into account in the design of the system

Compliance with the principle of minimisation

Summary of contributions

Contributors note that the **principle of minimisation** seems difficult to articulate with the development of an AI system.

In particular, contributors report that it may be difficult **to anticipate the most appropriate architecture** before testing the system on data, for example in the pilot phase.

CNIL' response

While the minimisation principle requires that the method chosen to achieve an objective is as data-efficient as possible, it does not provide for **an explicit threshold** and does not prohibit the collection of large datasets. On the other hand, the CNIL calls for the best possible anticipation of data collection and identification of the necessary data before embarking on the collection, in order to allow processing only what is **strictly necessary** for the design of the AI system. The conditions for compliance with the principle of minimisation have been clarified in this regard: deep learning methods must be reserved for cases where **no more economical alternatives exist,** and must be justified. Similarly, solutions requiring the use of particularly identifying types of data such as videos or photos must be necessary to achieve the objective.

On the role of the ethics committee

Summary of contributions

Several contributions requested clarification on the **role of an ethics committee.**

CNIL' response

The establishment and consultation of such a committee is a **good practice** to be associated with the validation of an AI development project that cannot be assimilated to governance, nor replace the competent authorities in the field. Although it may include external members, its role is to give an internal opinion on the relevance of an AI project.

Since the constitution of an ethics committee depends on the size and means of the structure on which it depends. An alternative good practice may also be to consult or appoint an "AI referent".

how-to sheet 7 – Taking data protection into account in data collection and management

Retention of data for audit purposes

Summary of contributions

The contributions pointed out that limiting the retention period for training data could be **an obstacle to carrying out audits of the AI system and in particular to measuring biases.**

CNIL' response

From a technical point of view, these analyses are often facilitated by access to training data. The possibility of conducting such audits is **of crucial importance to the security and safety of systems in the deployment phase,** in particular in view of the risk of discrimination they entail.

If it was already stated in the how-to sheet that audits planned in the maintenance phase may justify the retention of data, the CNIL wished to clarify its position. The sheet has thus been amended to:

- specify **that it is possible, after a sorting phase, to keep the training data for audit purposes;**
- indicate that **this retention requires the implementation of certain security guarantees** relating in particular to the restriction of access to data, their encryption and their pseudonymisation or anonymisation as soon as possible.

On the other hand, the retention of data cannot be justified in advance for the reason that a possible dispute requires access to the data for its resolution.