How-to sheets on the development of artificial intelligence systems (second series)

Public consultation form

The CNIL invites stakeholders to pool, if possible, their comments within a single contribution by integrating the various internal feedbacks or through their professional federation.

The contributions transmitted to the CNIL in this context are not public and their confidentiality will be ensured. Note, however, that these can be summarised and returned in the context of a possible synthesis of the public consultation published by the CNIL at the end of it.

Indeed, all contributions received by the CNIL can be the subject of a request for access as administrative documents (code des relations entre le public et l'administration).

In your contribution, report any element protected by literary or artistic property rights (specify, in this case, whether or not you allow it to be communicated), or by business secrecy.

Note that the CNIL is not required to follow your assessment on what is protected or not.

To send the CNIL this completed form, please send it by email to the following address: **ia@cnil.fr**

Most common formats are accepted (.pdf,.doc,.docx,.xls,.odt,.ods, etc.).

Questions asked in the context of the public consultation

Name: Kathleen McGrath

First name: Kathleen

(Mandatory) function: Data and Privacy Policy Manager

E-mail address (required): mcgrathk@huntonak.com

Name of organisation: Centre for Information Policy Leadership (CIPL)

Type of organisation (mandatory):

- **Independent think tank**

The Centre for Information Policy Leadership (CIPL) welcomes the opportunity to respond to the Commission Nationale de l'Informatique et des Libertés (CNIL) Consultation regarding how organisations develop AI models. For more than 20 years, CIPL has been a thought leader on organisational accountability and a risk-based approach as key building blocks of smart regulation, responsible governance, and use of data, as well as accountable development and deployment of artificial intelligence (AI). CIPL's *Ten Recommendations for Global Regulation*[1] proposes a layered, three-tiered approach to AI regulation that would protect fundamental human rights and minimise the potential risks of harm to both individuals and society while enabling the responsible development and deployment of AI. Our recent report, *Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework*[2], evidences best practices and case studies on how 20 leading organisations are responsibly developing and deploying AI through the lens of CIPL's Accountability Framework. Based on CIPL's independent research and observations, we provide input to the CNIL's public consultation below.

***

AI is the quintessential "rapid technological development" that is "transform[ing] both the economy and social life" as described by Recitals 6 and 7 of the GDPR. AI offers immense benefits and opportunities but also creates challenges and risks. As recital 4 GDPR sets out, the rights of the individual to the protection of their data must be carefully weighed against other fundamental rights. CIPL has long emphasized that regulators must evolve the interpretation of GDPR principles in light of technological evolution to ensure they remain valid and fit for purpose.[3] CIPL greatly appreciates the CNIL's leadership in advancing practicable and rights-preserving approaches to technological developments.

How-to sheets on the creation of datasets for the development of AI systems	**Comments and suggestions**
Legal basis for legitimate interest and development of AI systems	• CIPL supports the CNIL's notion that legitimate interests, rather than consent, would be the appropriate legal basis for the development of AI systems. • The legitimate interests legal basis relies on a balancing test that involves assessing the risks relating to the data processing activities and defining measures to mitigate these risks. It is effectively grounded in risk-based organisational accountability, for which CIPL has long advocated. It enables the processing of personal data when it does not

[1] CIPL White Paper - Ten Recommendations for Global AI Regulation (English) October 4, 2023 https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ten_recommendations_global_ai_regulation_oct2023.pdf

[2] CIPL AI Third Report - Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework February 21, 2024 https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_building_accountable_ai_programs_23_feb_2024.pdf

[3] CIPL/Hunton Andrews Kurth Legal Note - How the GDPR Regulates AI
March 12, 2020 https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-hunton_andrews_kurth_legal_note_-_how_gdpr_regulates_ai__12_march_2020_.pdf

result in a risk of harm to the interests and fundamental rights and freedoms of individuals that overrides the benefits of the processing. It also promotes the protection of individuals, as it requires organisations to undertake the necessary risk assessments, define the mitigation measures, train employees on risks and mitigation measures, monitor the continued effectiveness of the mitigations, identify potential compliance gaps, fix them, and keep improving the level of protection. Therefore, the flexibility provided by the legitimate interests legal basis, coupled with organisational accountability and its inherent risk-based approach, makes the legitimate interests basis a key enabler of responsible AI.

- CIPL furthermore strongly agrees with the CNIL's assertion that legitimate interests should be understood broadly and that commercial interests may also amount to a valid interest. Recital 47 of the GDPR provides some examples of the types of cases in which organisations may be able to rely on the legitimate interests legal basis, including prevention of fraud or direct marketing. However, the list is non-exhaustive and there are other common business practices which are generally considered to be a valid interest, such as improvement of services and cybersecurity measures.[4] The CNIL rightfully adds to the list. CIPL would support adding the development of AI systems as an example of a legitimate interest.
- As discussed above, an organisation relying on legitimate interests to develop or deploy AI is required to conduct a balancing test to assess whether individuals' interests override the interest being pursued, which effectively entails performing a risk assessment on the proposed processing activity. We welcome the CNIL's acknowledgement that an organisation's commercial interest may converge with the interest of the public or wider society and that this can potentially give more "weight" to the organisation's interests when carrying out a balancing test.
- CIPL firmly believes that legitimate interests can also include societal benefits, and the potential benefits of AI appear to be large across the economy and society in fields as diverse as medicine, science, agriculture and business. Consistent with the 2013 opinion of Advocate General Jääskinen[5] in the context of search engines, which informed the landmark CJEU Costeja decision, it is also appropriate for controllers developing AI models to rely on legitimate interests as their legal basis for processing personal data. AI brings broad capabilities, such as search, language, vision, reasoning or human interaction which can serve as the base for use-specific applications, bringing broad benefits to information access, dissemination, and the advancement of new technologies. The reasoning potentially holds even more true for AI training models,

[4] See a list of case studies in CIPL White Paper - How the "Legitimate Interests" Ground for Processing Enables Responsible Data Use and Innovation, July 1 2021 https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_white_paper_-_how_the_legitimate_interests_ground_for_processing_enables_responsible_data_use_and_innovation_1_july_2021_.pdf

[5] Opinion of Advocate General Jääskinen, https://curia.europa.eu/juris/document/document.jsf?docid=138782&doclang=EN, para 95. To quote from this paragraph: "As to the criteria relating making data processing legitimate in the absence of a data subject's consent (Article 7(a) of the Directive), it seems obvious that provision of internet search engine services pursues as such legitimate interests (Article 7(f) of the Directive), namely (i) making information more easily accessible for internet users; (ii) rendering dissemination of the information uploaded on the internet more effective; and (iii) enabling various information society services supplied by the internet search engine service provider that are ancillary to the search engine, such as the provision of keyword advertising. These three purposes relate respectively to three fundamentals rights protected by the Charter, namely freedom of information and freedom of expression (both in Article 11) and freedom to conduct a business (Article 16). Hence, an internet search engine service provider pursues legitimate interests, within the meaning of Article 7(f) of the Directive, when he processes data made available on the internet, including personal data."

which work with both structured and unstructured data, while search engine data is indexed and thus somewhat structured. Furthermore, we submit that, while the risks to be considered may vary, the legitimacy of the interest, and thus the appropriateness of the legal basis, does not change depending on whether or not the data was collected directly from the individual or from other sources.

- CIPL notes that beyond the right to privacy, other fundamental rights of individuals should be considered. For instance, insufficiently diverse data sets lead to biased outputs of AI models with a discriminatory impact on individuals or groups of individuals. AI models benefit from being trained on an extensive range of data to be useful for deployment in a wide range of contexts, for example in the diverse fields of education and research. Exposure to a wide and diverse dataset also benefits society by reducing the risk of inaccurate, biased, or even harmful outputs. It is critical that controllers have in place demonstrable safeguards to protect all these fundamental rights and mitigate risks (e.g., performing risk assessments and DPIAs, ensuring data quality, redress, and providing appropriate transparency).
- CIPL agrees with the CNIL that controllers must ensure not just the appropriate legal basis, but compliance with all the other provisions of the GDPR, such as data security, transparency, and rights of individuals, as well as non-infringement of intellectual property, product safety, and other laws.
- It is also important for the CNIL to recognise the distinct roles played by developers and deployers of AI systems with respect to risk assessment. Developers should be required to account for their role in the building and training of AI models and reasonably foreseeable uses, and deployers should be responsible for assessing risks associated with their particular deployments of the systems.
- AI system developers, providers, and deployers should also be incentivised to use privacy-enhancing and privacy-preserving technologies where feasible and appropriate, such as synthetic data, differential privacy, and federated learning.[6] The questions of how much data is necessary during the training and development stages and how best to implement data minimization measures in these contexts, including PETs, are complex ones and must be considered carefully. While emerging mitigation measures, such as synthetic data, hold promise for lowering the dependence on personal data, unduly limiting access to data or over-relying on such data minimising methods risks creating negative impacts on the development of models and hindering efforts to prevent and mitigate unintended bias (see also above).[7] Data minimization does not mean that only small volumes of data can be used in model training. Rather, data minimization in this context can be interpreted to require an appropriate balancing that reduces the amount of personal data used to what is necessary across the lifecycle of an AI system, permitting the development of a high-quality model and user experience. Stated differently, "data minimization" cannot mean using less data than would be necessary and appropriate to ensure the particular AI model's high quality.

[6] For more information on PETs, please see CIPL, Privacy-enhancing and Privacy-preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, December 2023, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-understanding-pets-and-ppts-dec2023.pdf.

[7] See CIPL Second AI Report: Hard Issues and Practical Solutions, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_second_ai_report_-_hard_issues_and_practical_solutions_01.17.2020.pdf and CIPL White Paper - Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age, December 12 , 2023 https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl-understanding-pets-and-ppts-dec2023.pdf

Legitimate interest: focus on open-sourcing models	N/A
Legitimate interest: focus on web-scraping	• As outlined above, legitimate interest is a valid, and appropriate, legal basis for scraping publicly available personal data and processing this data for the purpose of training Generative AI (Gen AI) models.[8] • Web-scraping has been a common practice for the provision of digital services to users: for example, internet search engines use web crawlers to index content. Many developers now rely on web scraping of publicly available data for training Gen AI models. It is at the core of how many Gen AI models are trained, as the breadth and variety of data are foundational to model quality and functionality and there is significant media coverage and public discussion around the practice of using publicly available data to train AI models. • The CNIL recognises that data scraping has become a widespread practice for generative AI systems that use large amounts of freely accessible online data. We would welcome more clarity, however, on the definition of 'accessible online data', specifically in relation to publicly available data that is behind a paywall or where accessibility is subject to account settings or restrictions. As a point of departure, the CNIL's recommendations in the "Additional Safeguards" section of the consultation to "limit the collection to freely accessible data (i.e. content accessible to any unregistered user without account creation) that were manifestly made public by the data subjects" and to "provide data subjects with the option to object to the processing at their discretion" seem useful, but merit further consideration with respect to feasibility and potential impacts on model performance. • Regardless, controllers should put in place effective safeguards to ensure that personal data included in the collected data sets are processed responsibly. • Organisations must be deliberate when it comes to the use of personal (or otherwise protected) data for the purposes of AI model training and development. This includes implementing effective safeguards to limit the impact on individuals' rights and freedoms. Furthermore, the ingestion of data sources that are potentially rich in personal and sensitive data, in particular where data is ingested from publicly available sources, can be limited in some instances. Web crawlers may be set to avoid certain data sources that are rich in personal data: for example, in May 2023, OpenAI's chatbot ChatGPT published a way for EU residents to request to have their personal data removed from the training databases[9]. In some situations, further filtering can be done during the pre-training phase, via pattern recognition algorithms for instance, which can be especially critical when it comes to data of more vulnerable groups, like children. At the same time, it must be noted, that reliably identifying personal data in pre-training datasets can be technically challenging, given the scale and structure of the

[8] See CIPL, "Response by the Centre for Information Policy Leadership to the Information Commissioner's Office's Consultation on the Lawful Basis for Web Scraping to Train Generative AI Models", March 1, 2024, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_response_-_ico_consultation_on_the_lawful_basis_for_scraping_data_for_generative_ai__mar_2024_.pdf.

[9]

	datasets involved, and could require additional processing, with organisations having to identify personal data in datasets where the personal data was otherwise incidentally included (and was not sought out or used to identify data subjects). • In that context, CIPL supports the CNIL's position that the incidental collection of sensitive data should be accounted for and cannot be excluded, in line with the views of the CJEU in the context of search engines. Furthermore, regarding the deletion of data, the CNIL should recognize that identifying individuals in web-scraped data in response to deletion requests could require additional processing of personal data, which is explicitly addressed in Art. 11 GDPR (i.e., processing which does not require identification). Organisations may also be able to show here that the required effort for an AI system developer to identify and subsequently provide the relevant privacy information to each individual whose data may have been collected meets the standard of disproportionate effort under Article 14 (5)(b) GDPR.[10] • In addition to being mindful of the provenance of the data used for AI training, organisations must also monitor and verify their AI systems, including, for instance, through adversarial testing like red teaming. Organisations should take appropriate measures to ensure guardrails exist around the output (see further below), commensurate with their roles and responsibilities in the AI value chain. It is important to point out that this is not just for the AI developer, but also the deployer or other third party that implements the model into an AI system. In many cases, including with open-source models, the developer does not determine the purposes and means of processing personal data for the deployment and does not have access to the personal data the deployer and user processes – for example, they are not in control of the input and output at this stage of the AI system.
Informing data subjects	• CIPL agrees with the CNIL's position that transparency is key to educating individuals on how an AI system uses personal data throughout the AI lifecycle. Organisations should make it possible for individuals to understand how their personal data is being used, and transparency measures should enable users to exercise their privacy rights where possible and when appropriate (e.g., right to object to the processing of their personal data, the right to restrict its processing, and the right to obtain its rectification or erasure), as well as helping users to understand the privacy settings they can utilize. This creates trust in the organisations' handling of data and enables individuals to seek redress where necessary. • However, we also consider the ability for organisations to satisfy transparency requirements to be dependent on context, including the purpose and intended use of the model, the organisations' relationship to the end user, and the intended audience for the information. The level of detail provided by transparency measures must be sufficiently tailored to the audience and must be proportionate to the risk posed by the processing. • In this context, CIPL commends the CNIL for recognizing that providing the suggested level of detail regarding data collection, such as specific information on the sources, types and categories of personal data used to develop the model, as well as the purpose(s) of processing, recipients, retention period, and rights of the data subject, may in some cases be impossible, not fully determined yet, or require a disproportionate effort. Such disclosures could potentially also require organisations to

[10] Please see CIPL's response to the UK Information Commissioner's Office consultation on Engineering Individual Rights into Generative AI Models, June 20, 2024, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/ci-pls_response_to_4thico_gen_ai_consultation_on_engineering_individual_rights_into_generative_ai_models.pdf.

identify personal data and link it to specific individuals where doing so would not have otherwise been necessary (or even possible). While it may theoretically be possible to do so in more common data structures, in the vast training data sets at the heart of LLMs there may be circumstances where data are impossible to conclusively link to individuals or groups of individuals (e.g., identifying one "Antoine Martin" from all the "Antoine Martin's" online). In many cases, training data is unstructured and not organised by reference to individual identifiers or in a way that facilitates querying the data by reference to particular identifiers. It would be helpful for the CNIL to clarify that, in such instances, developers can meet transparency and notice requirements through publicly accessible notices more generally explaining that publicly available data is being used, as the CNIL suggests in its draft How-to Sheet for this topic.

- Where data is directly collected from individuals, i.e., separate from web-scraping for instance, and a direct link is established between the controller and the individual, organisations should determine which transparency measures will be most effective to communicate how their data is used so that individuals can meaningfully exercise their rights. However, exemptions should apply where data is not collected directly from individuals (such as in cases where data is collected via web-scraping). As discussed above, identifying individuals for notification purposes in web-scraped data would require large-scale additional processing purely for notification purposes, as addressed in Article 11 GDPR (i.e., processing which does not require identification), or in case of direct collection, it may require updating notices. Organisations may be able to show that the required effort for a model developer to identify and provide the relevant privacy information to each individual whose data may have been collected meets the standard of disproportionate effort under Article 14 (5)(b) GDPR. CIPL agrees with the CNIL that in assessing whether the effort is disproportionate in nature, organisations should weigh the impacts on the privacy of individuals on one hand and the efforts required to individually inform on the other. We support the notion that this should include whether the organisation has put any safeguards in place or implemented any technical and organisational security measures to protect individuals' rights.
- To that end, CIPL strongly supports the continued development, adoption, and implementation of privacy-enhancing and privacy-preserving technologies (PETs/PPTs) in the context of the entire AI lifecycle. These tools can further minimize the risk of identifying an individual through their personal data in the context of AI. For example, synthetic data may eventually be able to supplement real-world data during model training, differential privacy could be used to add noise for certain training sets to ensure individuals whose data is present in the training data cannot be explicitly or implicitly identified, and homomorphic encryption can keep data secure during training by keeping data encrypted throughout the entire process. PETs/PPTs can and should be combined, meaning that multiple techniques can be used alongside each other to further strengthen protections. For instance, use of filtering personal data, differential privacy, or k-anonymization can reduce the risk of memorization, and application of filtering to model outputs to mitigate the risk of inferred (or hallucinated) data outputs that contain sensitive information. CIPL outlines how PETs support data protection principles in our report *Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age*. We encourage the CNIL to continue their support of the use of PETs and PPTs, and to bring more clarity to the sector players about the circumstances in which the use of such technologies is appropriate or recommended.[11]

[11] See CIPL's PETs paper, supra n 8

	• In the context of Gen AI chatbots that allow user prompts, individuals should be informed where such prompt data is used for model training. This can happen through a number of methods, such as privacy notices, legal terms, and just-in-time prompts, to ensure the user remains in control of the data they provide to the system. Some organisations may also limit "chat memory" or offer users the ability to prompt the chatbot to remember or delete its memory of certain data they have put into the chat. Other input filters, including de-duplication practices, can also be useful to minimize personal data in a training set.[12] • CIPL agrees with the CNIL that data controllers should indicate clearly when they are not able to identify individuals, including for purposes of exercising their rights. In such circumstances, information can still meaningfully be provided by different methods and at different appropriate points throughout the lifecycle of the data (e.g., in publicly accessible privacy notices or other disclosures), independently of the identifiability of an individual. Transparency and notice requirements can be met through public disclosures and information campaigns, accessible privacy notices, or other informational resources explaining how data is used in the context of the model, for example.[13] We encourage the CNIL to retain these points and make them clear with practical examples or scenarios in its final guidance. • As a general rule, CIPL believes that the responsibility to inform individuals about the use of their data must fall to the entity closest to the individual from whom the data is collected, whether that be during development or deployment. For example, a partner of the developer who provides personal data to the developer for training purposes is closer to the individual than the developer. We encourage the CNIL to acknowledge this point on the responsibilities of deployers, particularly in the context of the right of access; deployers should be responsible for complying with access requests received in relation to personal data they process during their particular use of the AI. Similarly, developers may have greater responsibility than deployers to provide transparency about the source of the data used during AI training. • CIPL welcomes the CNIL's acknowledgement of Article 53 EU AI Act in this context as a source of information. Consistent application of overlapping legal provisions is key to legal certainty and successful uptake of AI as envisioned by Article 1 EU AI Act. AI developers can share information about their models with deployers through model documentation (e.g., model or system cards), to allow deployers to inform individuals accordingly. Transparency through such tools should extend beyond information about the data used, the performance, use cases, capabilities and limitations of the model. In all instances, the level of transparency should be balanced not only with the need to protect intellectual property rights, copyright, confidential information, and trade secrets, but also the vulnerabilities of AI systems and the potential net societal benefit that may outweigh individuals' rights. Transparency should not come at the expense of other important factors, such as usability, functionality, and security of the system, or create additional burdens for users.

[12] Please see below in section 4.1 where we go into more detail on the availability of output and input filters and their effectiveness at minimizing personal data processing.

[13] CIPL recognizes that many data protection authorities have published their own guidelines on generative AI, many of which address the topic of data subject rights. Of note, the EDPB's ChatGPT task force's recent report states that "in line with Art. 25(1) GDPR, the controller shall…implement appropriate measures designed to implement data protection principles in an effective manner and to integrate the necessary safeguards into the processing" to meet GDPR requirements and protect data subject rights. Another guideline from the German Data Protection Conference reminds organisations to ensure that data subjects can exercise their rights, such as their right to erasure and rectification, through the appropriate technical and organisational measures.

Respecting and facilitating the exercise of data subjects' rights	• In the context of exercising individual rights, the CNIL makes an important distinction between the exercise of data subject rights on the training data set versus the AI model itself. CIPL supports this separation into the different modules and components of an AI system, which is also followed by the recent discussion paper on personal data in LLMs issued by the Hamburg DPA.[14] Whether or not there is a sufficient link to any individual, and the exercise of their rights, is driven by context, type of application and the module of AI in question. We welcome this differentiation. 4.1 User Rights in the context of training data • CIPL welcomes the CNIL's general acknowledgement that controllers do not need to identify individuals whose data may be contained in a training data set, or to store or collect additional information solely for the exercise of user rights on the training datasets in accordance with Article 11 GDPR. A direct link to an individual is generally not necessary for intended training and it would be counterintuitive. Creating identifiability in the hands of a controller where it did not exist before would be contrary to Art. 11(1) GDPR. However, the CNIL also suggests creating technical and organisational measures and providing additional data to data subjects at their request. For example, the CNIL proposes that controllers undertaking web-scraping retain domain names and URLs of the web pages on which the data was collected in order to be able to identify an individual for notification purposes, contrary to Arts. 5(1)I[1][15] and 11(1) GDPR if the controller does not need to retain the domain names and URLs otherwise. At the same time, the CNIL emphasizes that "the use of overly intrusive methods cannot be justified." Requiring developers to retain the domain names and URLs of web pages on which data is collected for the purposes of identifying data subjects would ultimately equate to cataloguing the internet; in some cases, doing so would involve the retention of millions of domain names and URLs. It would be helpful for the CNIL to provide more clarity on the concept of "intrusive methods" as well as examples.[16] • In addition, CIPL cautions that maintaining and monitoring a "push-back list" as suggested by the CNIL for model training could be technically challenging and could, paradoxically, increase privacy risks by requiring the identification of individuals in data sets where this would not otherwise have occurred. • Concerning user rights in the context of AI, CIPL submits that, as a general rule, it is important to consider the extent to which the societal benefit of processing data for the purpose of further developing a model, which requires sufficiently diverse and "good" data sets, may outweigh the risks to individuals. If certain representative groups

[14] Hamburg Data Protection Authority, Discussion Paper: Large Language Models and Personal Data, July 2024 https://datenschutz-hamburg.de/fileadmin/user_upload/HmbBfDI/Datenschutz/Informationen/240715_Diskussionspapier_HmbBfDI_KI_Modelle.pdf

[15] Article 5(1)(c) of the GDPR provides that "[p]ersonal data shall be … limited to what is necessary in relation to the purposes for which they are processed."

[16] Moreover, information regarding the composition of training datasets will not meaningfully help deployers make determinations regarding a model's fitness for their particular use case. The fact that a model has been trained on certain data does not mean it will perform as needed in a specific use case. For example, even if a deployer has access to information that a model was trained on data across a variety of languages, including English, Mandarin, Hindi, Spanish, French, in order to assess how well the model performs in transcribing Mandarin in a particular application, the deployer must test the model in order to confirm its performance. As such, for general purpose models, model developers should only be expected to provide information about the performance, capabilities, and limitations of the model.

of individuals were to disproportionately have their data repressed from AI training data by way of objecting in accordance with Article 21 GDPR, for instance, this could have a negative impact on the entire model by introducing bias and thus potentially diminish its utility for society at large.

- A useful analogy can be drawn from the rules around automated decision-making (ADM) under Article 22 GDPR, whereby an individual has the "right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her". This is an example of having to clear a certain risk/harm threshold before a particular data protection right can vest. While it would be difficult to compile a list of examples that would conclusively establish what constitutes an "impactful" automated decision for the purposes of Article 22, asking relevant questions in a specific context will provide factors that may assist in making this determination. Similarly, where an individual objects to the processing of their data under Article 21 GDPR (i.e., the right to object) in the context of AI models, CIPL believes the subsequent analysis of compelling legitimate grounds for processing must then take into account the potential societal benefits of the particular model, such as the need for it to be built on representative data, and to weigh those against the risks associated with processing the data. In this context, synthetic data might play a role in ensuring sufficiently diverse and balanced data sets.
- It may also be helpful for the CNIL to explicitly acknowledge within their guidance that there may be special circumstances where organisations are unable to comply with erasure/rectification requests because the associated data are subject to data retention requirements, including data that are in conflict with data retention requirements from other legal acts, such as anti-money laundering requirements, are under a litigation hold, or are collected and processed in very specific context (such as medial trials), and thus, prohibited from being further processed, including deletion or modification of the data. This is particularly relevant for organisations operating in financial services and bio-pharmaceuticals but may also be important for other industries.
- CIPL further strongly supports the CNIL's acknowledgement that the exercise of user rights must respect the rights of third parties, including data protection rights, intellectual property rights, and trade secrets.

<u>4.2 User Rights in the context of the AI model</u>

- The CNIL submits that training an AI model with personal data might in some cases lead to the application of the GDPR on the model in question. Whether models can be considered to contain personal data or not, it will be important to distinguish between the model itself and the output of the model. The CNIL is quite clear that cases where individuals are identifiable in a model are the exception. In the cases of LLMs, the ingested data is broken into smaller tokens, sometimes below the level of a single word. Within the model, "Antoine Martin" may not exist as such, and the output is not necessarily a direct reflection of the input, but may be something (newly) generated, based on statistical probability calculated by the model.[17]
- It is therefore of great importance to be transparent about the limitations of a model I the deployer, but also the end user. Deployers must be put in a position to understand

[17] See also the excellent description in the Discussion Paper of the Hamburg DPA, supra n 15

the appropriate use cases of the model (for instance, through model cards or terms of use) to ensure they are taking appropriate steps to safeguard user rights. End users interacting directly with a model, such as via a chatbot, must be made aware in clear and understandable language what the limitations of the output of the model are, especially with respect to the accuracy of the output. This can include labels or watermarks but may also take the form of prompts to verify output with additional sources, for example.

- CIPL supports the verification of AI models, through red teaming exercises and fine-tuning of outputs as an accountability measure[18] as well as providing feedback loops for continued evaluation. As the technology of these verification techniques for AI models continues to evolve, CIPL would also like to emphasize the importance of incentivizing further investment into continued research to foster better understanding of their efficacy.
- With respect to the rights of erasure or rectification on the model itself, CIPL agrees with the CNIL that at present the state of the art in technology is not advanced enough to allow identification of personal data from the weights of AI models alone. Research on "machine unlearning", which may eventually enable the deletion of specific points of data or eliminate their impact on AI model outputs without losing the "learning" from them, is promising. However, it continues to encounter a number of challenges. Machine unlearning remains resource intensive, and it can affect the performance of the model, depending on the unlearning method used and the importance of the removed data (i.e., removing data that had a significant impact on the model's learning might degrade or impact the model's performance).
- As the research makes progress, CIPL encourages the CNIL to consider guidance on the acceptable level of deletion efficacy (i.e., a model's ability to remove specific data or knowledge from a model in an effective and irreversible way) that could satisfy data erasure requests. As there are different mathematical levels of unlearning in a model, regulatory guidance could provide clarity around what would be considered an acceptable level of "unlearning". It will also be important to consider the possible "unintended consequences" when determining an acceptable "cut-off", particularly for models that deploy differential privacy techniques. Differential privacy requires a high-density data set with added noise to render single individuals unidentifiable. However, honoring deletion requests at scale or mandating proof of deletion may undermine the very privacy protections that such techniques provide.
- While machine unlearning cannot yet be considered sufficiently well developed for widespread deployment, output filters in the context of generative AI models can be an effective mechanism to address data subject rights. In practice, these can be commonly understood to be processes by which inputs (such as prompts) and outputs are screened to detect personal data and trigger associated actions. This involves blocking future model responses related to an individuals' information so that the learning from the training remains, but the objection is applied moving forward. For example, if Antoine Martin requests that his data no longer be used to provide outputs in a model, some organisations will implement a blanket "block" on data related to all "Antoine Martins" rather than just the one individual and could harm the performance of the model.
- However, the CNIL appears to suggest that filtering the outputs of the system to enable rights requests would be less relevant than re-training the model. CIPL submits that, barring exceptional circumstances, retraining a full model in each individual instance of a request where training data may still be available, could be disproportionately

[18] See CIPL's AI Third Report, supra n 2

	burdensome for the organisation to conduct business, and also wider societal impacts.[19] Frequent re-training may negatively impact the learning of the model and by extension the utility for the broader society, where certain data is no longer present in the learning. • As a general point, CIPL strongly supports the CNIL's emphasis on derogations to the exercise of rights on datasets or on the AI model. The individual's rights to their data are qualified rights and must be balanced against other factors. Therefore, there must be limitations, such as when the request would require disproportionate efforts. It is also critical that there be allowance for circumstances in which requests can be refused, such as where they are clearly vexatious or excessive under Article 12 GDPR. For example, there are instances when malicious actors may be exercising abusive data subject access rights with the sole purpose of gathering information so they can bypass a cybersecurity or fraud prevention system. In this case, the social good to ensure cybersecurity or fraud prevention could take precedence over the individual right.
Annotating data	N/A
Ensuring the safe development of an AI system	• CIPL proposed early on that the application of AI could require a Data Protection Impact Assessment (DPIA) based on the criteria of Article 35 GDPR. [20]. We agree with the CNIL that the DPIA is a useful tool to identify risks and design appropriate measures. A DPIA must contain a systematic description of the proposed processing, its purpose and the legitimate interest pursued, as well as an assessment of its necessity and proportionality, its risks, and the measures envisaged to address those risks. • This captures the very essence of CIPL's approach to accountability and reflects the overarching principle: the higher the potential risk of harms, the more organisations must do to demonstrate the elements of organisational accountability. This includes AI governance programs and controls that are continuously adapted and calibrated in keeping with technological and regulatory developments. A well-developed, comprehensive accountability framework or program provides organisations with the tools and processes needed to implement relevant legal requirements and standards, as well as internal ethics standards and other internal best practice goals. CIPL has issued a report mapping emerging best practices for accountable AI programs to the CIPL accountability framework which aligns with a number of the CNIL's proposed measures. We are attaching the report to this submission.[21] • CIPL appreciates the numerous practical examples and recommendations for the assessment and mitigation of risk specific to different phases of an AI system that the CNIL is providing. It is important that this list remains indicative and will not preclude the development of alternative approaches that would achieve the same goal of ensuring the security of AI systems. In our research, CIPL has found that accountable organisations have already adopted many of the proposed or similar measures for their AI governance, such as:

[19] We may also have to consider that the frequent retraining of an AI model based on individual requests, especially when considering the timelines of the GDPR as the CNIL suggests, may ultimately have a negative impact on the environment given the often immense computing power required for model (re)training.
[20] See CIPL/Hunton Andrews Kurth Legal Note - How the GDPR Regulates AI, supra n 3
[21] See CIPL's Third Report AI, supra n 2

- Offering AI training and education modules: Organisations are upskilling their workforces by offering training courses and encouraging enrollment in external certifications and programs
- Investing in multidisciplinary and diverse teams: Organisations find that AI governance teams must be cross-disciplinary to be effective in terms of expertise (e.g., technology, social, industry, legal, human resources, government relations, ethics) and diverse across multiple factors, including demographic (e.g., gender, age, religion, sexual orientation, and ethnicity), and geography
- Creating additional oversight bodies, such as AI ethics committees either internally or externally
- Incorporating red-teaming or adversarial testing for their AI models into the governance framework

- With respect to the recommended security measure of "providing for the possibility of stopping the system," which is positioned as mandatory for uses that fall within the scope of automated decision-making with a legal or similarly significant effect under Article 22 GDPR, the CNIL should clarify the extent to which it applies on deployers vs. developers, taking into account the extent of their respective control over the use of the application. In many circumstances, the developer will not be in a position to even know the ultimate use of the model and whether it falls within the scope of Article 22 GDPR.
- To the extent that the recommendations overlap with obligations or measures introduced in parallel legislations (i.e. the Digital Services Act, EU AI Act), CIPL recommends close regulatory cooperation to ensure a unified approach and common interpretation of similar concepts. Finally, CIPL commends the CNIL for underscoring the importance of sound security measures, such as verifying the reliability of training data, putting in place access controls, applying pseudonymization or anonymization where appropriate, and undertaking technical measures such as encryption, and using privacy-enhancing technologies (PETs) such as synthetic data.[22]

[22] See CIPL's PETs paper, supra n 8