CIPL
Centre for Information Policy Leadership
HUNTON ANDREWS KURTH

Response by the Centre for Information Policy Leadership to the Information Commissioner's Office's Fourth Consultation on Engineering Individual Rights into Generative AI Models

Submitted June 20, 2024

The Centre for Information Policy Leadership (CIPL) welcomes the opportunity to respond to the Information Commissioner's Office's (ICO) Consultation regarding how organisations developing generative AI (genAI) models can help individuals exercise their individual rights. 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 ICO public consultation below:

1. **Do you agree with the analysis presented in this document?**

   - CIPL agrees with the ICO's position that where personal data is collected directly from individuals, the organisations collecting it must explain to individuals how their data will be used and how they can exercise their data subject rights. Transparency is key to educating individuals on how the AI system uses personal data throughout the AI lifecycle and to put individuals in a position to exercise their rights where appropriate. In this context, we specifically support the ICO's analysis 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 client of the developer who provides personal data to the developer for training purposes is closer to the individual than the developer. We encourage the ICO to acknowledge this point in its future guidance 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.

   - However, some exceptions to the right to be informed should apply where data is not collected directly from individuals (e.g., in cases where data is collected via web-scraping). Data sets used for training genAI are vast, generally unstructured, and may include personal data only incidentally. Identifying individuals for notification purposes in web-scraped data

[1] CIPL, "Ten Recommendations for Global AI Regulation", October 2023, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_ten_recommendations_global_ai_regulation_oct2023.pdf.

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

would require large-scale additional processing purely for notification purposes, which is explicitly addressed in Art. 11 UK GDPR (i.e., processing which does not require identification). The effort of the organisation in informing the individual must in each case be contextually balanced against the harms to the individual's rights. Organisations would also likely be able to show that the required effort on a model developer's part to attempt to identify and subsequently provide the relevant privacy information to each individual whose data has been collected through web scraping meets the standard of disproportionate effort under Art. 14 (5)b UK GDPR. Furthermore, while it may be possible to identify certain individuals in more common data structures, in these vast training data sets, some data may be difficult or impossible to conclusively link to individuals (e.g., identifying one "John Smith" from all the "John Smith"s online). Also, the quality and accuracy of web-scraped datasets may not be up to date. Instead, we agree that 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.[3] We encourage the ICO to retain these points and make them clear with practical examples or scenarios in its final guidance.

- Where AI developers are unable to identify individuals associated with personal data contained in their training datasets, they should be able to rely on the exception set out in Art.11(2) UK GDPR. Additionally, the removal of data could lead to a degradation of the quality or representativeness of the model (please see additional discussion below). Data subject requests must therefore be handled on a case-by-case basis, in accordance with Art. 11(2) UK GDPR. We encourage the ICO to make this clear with practical examples or scenarios in its final guidance.

- The ICO's consultation also states that "for web-scraped datasets, the processing of personal data to develop genAI models is likely to be beyond people's reasonable expectations at the time they provided data to a website."[4] We encourage the ICO to reconsider this statement in its guidance as web-scraping is in fact a common practice that enables many features of the internet (for example, indexing used by search engines).

- The ICO makes it clear in their analysis that transparency to individuals regarding an organisation's collection, processing, and use of data is crucial. CIPL agrees and believes that organisations should make it possible for individuals to understand how their data is being

[3] CIPL recognises 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.

[4] See the ICO's fourth call for evidence: engineering individual rights into generative AI models, https://ico.org.uk/about-the-ico/what-we-do/our-work-on-artificial-intelligence/generative-ai-fourth-call-for-evidence/.

used and transparency measures should enable users to exercise their privacy rights where possible and appropriate and at the appropriate time (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). However, we recognise that the ability for organisations to satisfy such requests may be dependent on context, the purpose and intended use of the model, etc. Furthermore, the level of detail provided by transparency must also be proportionate to the risk posed by the processing, and organisations should recognise that the greater the risk posed by the processing, the higher the level of transparency that should be offered to individuals. Transparency also should not come at the expense of other important factors, such as usability, functionality, and security[5], or create additional burdens for users. Where data is directly collected from individuals, i.e., separate from web-scraping and a direct link is established between the controller and the individual, organisations must communicate how their data is used so that individuals can meaningfully exercise their rights where possible. In the case of data not directly collected from the individual, information should still 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).[6] AI developers can also share information about their genAI system with deployers through model documentation (e.g., model or system cards), thus allowing deployers to inform individuals about how their data was processed for the development of the relevant genAI system. None of these transparency obligations obviates the need for controllers to show a legal basis under the UK GDPR to the extent that personal data may be caught in scraped data, as CIPL also stated in our previous consultation response.[7] As an advocate of organisational accountability and taking a risk-based approach to data and AI regulation and compliance, CIPL believes that controllers should in all cases conduct an appropriate risk assessment to properly weigh the benefits

- CIPL also 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 minimise the risk of identifying an individual through their personal data in a given model. 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. CIPL outlines how PETs support data protection principles in our report, *Privacy-Enhancing*

[5] For instance, overly granular transparency might provide malicious actors with an “in” to the model that can ultimately compromise its security,

[6] See more on CIPL’s perspective on transparency in our GDPR Implementation Project’s “Recommendations for Implementing Transparency, Consent and Legitimate Interest under the GDPR”, May 2017, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_recommendations_on_transparency_consent_and_legitimate_interest_under_the_gdpr_-19_may_2017-c.pdf

[7] See more in CIPL’s Response to the ICO Consultation on the Lawful Basis for Web Scraping to Train Generative AI Models, March 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

and Privacy-Preserving Technologies: Understanding the Role of PETs and PPTs in the Digital Age[8]. We encourage the ICO to continue their strong support of the use of PETs and would, for instance, welcome the deployment of PETs in an AI context to be included in the ICO Regulatory Sandbox, given their significant potential.

- Particularly in the context of AI, it is also important to consider the extent to which the societal benefit of processing data for the purpose of further developing a model may outweigh the risks to individuals, such as ensuring sufficiently diverse and "good" data sets. A useful analogy can be drawn from the rules around automated decision-making (ADM) under Art. 22 UK 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".[9] This is an example of having to clear a certain risk/harm threshold before a particular data protection right can vest. As the ICO had previously stated, it would be difficult to compile a list of examples that would conclusively establish what constitutes an impactful automated decision for the purposes of Art. 22, and had suggested an alternative way to think about such impacts – by asking relevant questions in a specific context, and recognizing that certain factors may assist in making this determination.[10] Similarly here, where an individual objects to the processing of their data under Art. 21 UK GDPR (i.e., the right to object) in the context of genAI 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 against processing the individual's data. This will also play a role in the context of groups of individuals exercising their individual rights, which may impair fairness and statistical accuracy of the model, as pointed out by the ICO in their analysis. In this context, synthetic data might eventually play a role in ensuring sufficiently diverse and balanced data sets.

- Work on "machine unlearning" is currently underway. "Machine unlearning" is intended to enable the deletion of specific points of data or eliminate their impact on AI model outputs, which may ultimately support requests for erasure of data, for example. However, there are still many challenges to effective and efficient machine unlearning - it remains resource intensive, and it may 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

[8] 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

[9] UK GDPR, Article 22.

[10] CIPL's Response to UK Department for Digital, Culture, Media and Sport (DCMS) Policy Paper on Establishing a Pro-innovation Approach to Regulating AI, September 2022, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_response_to_uk_dcms_proposed_approach_to_regulating_ai_23_09_22.pdf

performance). Ongoing research aims to improve unlearning techniques to make them more practical and effective for model developers.[11]

- Additionally, CIPL would like to encourage the ICO to provide holistic and practical 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.

- We must also remember that tensions between individual rights and emerging technologies are not new. For example, the immutable nature of blockchain and distributed ledger means in principle that all transactions are recorded forever, and that deletion is not an option.[12] The French CNIL has acknowledged that it is technically impossible to grant a data subject's request for erasure when data is registered on a blockchain and that some level of identification is necessary part of the blockchain.[13] As a solution, the CNIL does encourage using hashes, cryptographic references, and other validators on-chain. This example demonstrates that privacy rights can be adapted to the realities of different emerging technologies with the aid of innovative technological solutions such as PETs. We encourage the ICO to consider technical limitations and practicalities in their ongoing consultation.

- It may also be helpful for the ICO 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, or are under hold due to litigation proceedings, and thus, prohibited from being further processed, including deletion or modification of the data. This is particularly relevant for organisations operating in financial services but may also be important for other industries.

2. **Where training or fine-tuning data is web scraped or collected in other ways, what measures do you think are effective to inform individuals about how, why and by whom their personal data is being processed?**
   - Please see our response to Question 1 above.

3. **What kind of information do individuals need in relation to their data in the context of generative AI so they can exercise their rights?**

[11] For a review of current research on machine unlearning, see Ken Ziyu Liu, "Machine Unlearning in 2024," April 2024, https://ai.stanford.edu/~kzliu/blog/unlearning.

[12] For more examples see CIPL Discussion Paper Digital Assets and Privacy, January 2023

[13] CNIL, "Blockchain and the GDPR: Solutions for a responsible use of the blockchain in the context of personal data", September 2018, https://www.cnil.fr/sites/cnil/files/atoms/files/blockchain_en.pdf.

- Organisations must be transparent about their data processing practices. This creates trust in the organisations handling of data and enables individuals to seek redress where necessary.

- The ICO's consultation states that specific information "on the sources, types and categories of personal data used to develop the model" should be made publicly available. Given the volume and unstructured nature of the data, it may not be feasible for developers to continuously or periodically review each individual data point nor to identify every type of personal data that could be contained in datasets. This would potentially also require organisations to identify personal data and link it to specific individuals, which would not have otherwise been necessary. It would therefore be helpful for the ICO to clarify that, in these cases, developers (i) are not expected to structure or index datasets for the purpose of providing information about their processing activities to specific individuals and (ii) can meet transparency and notice requirements through public notices generally explaining that publicly-available data is being used. 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 genAI systems and the potential net societal benefit that may outweigh individuals' rights. For example, there are instances when malicious actors may be exercising abusive data subject access 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 prevalence over the individual right.

- Where the ICO suggests that "the purposes for which personal data is being processed and the lawful basis for the processing" should be made publicly available, we would like to point to CIPL's previous consultation response, and encourage the ICO to acknowledge that developing and deploying a general-purpose model may constitute an example of a legitimate interest and can be a sufficiently specific purpose.[14]

- Especially in the context of genAI chatbots that allow user prompts, individuals must 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, just-in-time prompts, to ensure the user remains in control of the data they provide to the system. Where appropriate and possible, users should also have a way to request that their prompts not be included in model fine-tuning. Some organisations will also prevent their models from having too long a "chat memory" or offer users the ability to prompt the chatbot to remember or delete its memory of certain data they've inputted into the chat. As noted above, the responsibility to inform individuals about the use of their data must fall to the entity closest to the individual at the collection stage.

4. Are you aware of any innovative approaches to enabling data subject rights requests over training and fine-tuning data?

[14] See more in CIPL's Response to the ICO's 2nd Consultation on Purpose Limitation in the Generative AI Lifecycle, April 2024, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_response_-_ico_consultation_on_purpose_limitation_in_the_generative_ai_lifecycle__apr_2024_.pdf

- Some organisations use objection forms to provide individuals with an opportunity to object to processing. Specific genAI privacy notices and help hubs through accessible communications platforms in English or local languages as appropriate are also deployed alongside already existing privacy notices.

5. **What measures, if any, including input and output filters have proved effective in enabling data subject rights in the context of generative AI models?**
    - We agree that input and output filters are effective mechanisms to address data subject rights in this context and encourage the ICO to retain this point in its final guidance. While input/output filters are a straightforward concept in essence, for the purposes of guidance it would be helpful for the ICO to provide a clear and specific explanation of how they define such measures. 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.
    - Some organisations are utilising technical solutions to prevent data scraping from websites traditionally rich in personal data or behind paywalls and log-ins. A=Website owners may use the robots.txt file to provide directives for web crawlers on how the site should and can be crawled (e.g., what parts of the website they can and cannot crawl). Also, OpenAI's web crawler, GPTbot, for example, comes with an opt-out feature for website owners that can either disallow GPTBot from accessing the site entirely or to access only parts of the site.[15] However, it is important to note that the directives are advisory and website owners must rely on the compliance of the web crawler. Thus, AI developers using web crawlers to collect data should take care to read and respect the outlined directives.
    - Pattern recognition algorithms can be used in the pre-training phase to filter potentially private data out of any training data sets.
    - Organisations are also utilising output mitigation techniques, like output filters. This involves blocking future model responses related to and individuals' information so that the learning from the training remains, but the objection is applied moving forward. For example, if John Smith 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 "John Smith"s rather than just the one individual. This prevents the developer from having to trace the data back to that particular individual, which also may be technically difficult or impossible to practically do with third-party data.

[15] See more at https://platform.openai.com/docs/gptbot