Centre for Information Policy Leadership Responses to UK Information Commissioner's Office's Consultations on Generative AI and Data Protection

Submitted to the UK ICO 2024 and Compiled for APEC SOM3

The Centre for Information Policy Leadership (CIPL) welcomed the opportunity to respond to the Information Commissioner's Office's (ICO) consultations on Generative AI and Data Protection.

Enclosed here are CIPL's responses to three of those consultations:

- Consultation 1: The lawful basis for web scraping to train generative AI models
- Consultation 3: Purpose Limitation in the Generative AI Lifecycle
- Consultation 4: Engineering individual rights into generative AI models

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

Submitted March 1, 2024

The Centre for Information Policy Leadership (CIPL) welcomes the opportunity to respond to the Information Commissioner's Office's (ICO) Consultation on the lawful basis for web scraping to train Generative AI (Gen AI) models.

For more than 20 years, CIPL has been on the forefront of promoting organisational accountability and a risk-based approach as cornerstones of effective data protection law, policy, and oversight. As noted in our white paper *Ten Recommendations for Global AI Regulation*[1], CIPL advocates that any regulatory approach to AI should seek to protect fundamental human rights and minimise risks of harm to individuals and society, while enabling the responsible development and deployment of AI. CIPL has recently published a report entitled *Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework*[2], which showcases how 20 leading organisations are developing accountable AI programs and best practices for developing and deploying AI on the ground through the lens of CIPL's Accountability Framework.

CIPL's feedback on the Consultation will focus on 1) the ICO's analysis 2) the technical and organisational measures developers should implement to limit the ways customers can use Gen AI models, and 3) the aspects of this topic CIPL would like the ICO to consider in future publications.

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

CIPL is still shaping our thinking on public policy and governance of Gen AI given the fast evolution of the technology. At the same, we can draw on durable principles from our work to date in order to share perspectives for this consultation. CIPL previously assessed the application of GDPR to AI[3] and addressed key tensions in applying data protection principles to AI.[4] In the latter paper, we also emphasised the need for regulators to evolve the interpretation of GDPR principles in light of technological developments to ensure they remain valid and fit for purpose for Gen AI technologies. The ICO consultation seeks to do just that, and we welcome the approach.

CIPL agrees that legitimate interest can be a valid, and appropriate, legal basis for scraping publicly available personal data and processing this data for the purpose of training Gen AI models. Use of

---

[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.

[3] CIPL, "Artificial Intelligence and Data Protection: How the GDPR Regulates AI," March 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.

[4] CIPL, "Second Report: Hard Issues and Practical Solutions," February 2020, https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_hard_issues_practical_solutions_english_feb20.pdf.

legitimate interest as a legal basis for training Gen AI models requires putting in place appropriate guardrails to keep data safe, controls and accountability measures, as well as robust oversight and enforcement.

In order to rely on legitimate interest, a controller must be able to demonstrate that there is a specific valid interest to process personal data in this way. Similar to the 2013 opinion of Advocate General Jääskinen in the context of search engines, which informed the landmark CJEU *Costeja* decision,[5] it would similarly seem appropriate for controllers developing Gen AI models to rely on legitimate interest as a ground for processing personal data sourced from the internet for the purpose of training the models, given the broad societal benefits the technology can bring . The ICO has long supported a broad interpretation of what can constitute a legitimate interest, including societal benefits, and the potential benefits of Gen AI appear to be large across the economy and society in fields as diverse as medicine,[6] science, agriculture and business.[7] More specifically, in this case there is a legitimate interest to ensure robust, accurate, safe and non-biased functioning of algorithms by training models using large, diverse and high quality sets of data.

CIPL acknowledges that at present many developers rely on web scraping of publicly available data for training Gen AI models. Publicly available data is at the core of how many Gen AI models are trained; it is foundational to model quality and functionality. Gen AI requires data to learn how language incorporates concepts about relationships between people and the world. A Gen AI model is not so much a "database" holding data, but rather, an algorithm that has learned patterns and relationships in data and uses them to predict the next probable words or images in a sequence.

However, controllers must put in place demonstrable policies and procedures to ensure that personal data are processed responsibly. This includes instituting safeguards against models being used to harm the fundamental rights of individuals whose personal data may be processed within the models, such as by enabling users to derive inferences on specific people's sensitive characteristics.[8] Removing personal data from the data collection and training stage of Gen AI is an

---

[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."

[6] Andrew Myers, "Doctors Receptive to AI Collaboration in Simulated Clinical Case without Introducing Bias," https://hai.stanford.edu/news/doctors-receptive-ai-collaboration-simulated-clinical-case-without-introducing-bias.

[7] Adam Zewe, "Explained: Generative AI," *MIT News,* https://news.mit.edu/2023/explained-generative-ai-1109.

[8] *Google Spain SL and Google Inc v Agencia Espanola de Proteccion de Datos (AEPD) and Mario Costeja Gonzalez*, Case C-131/12, https://curia.europa.eu/juris/document/document_print.jsf?doclang=EN&docid=152065, at para 80. To quote from this paragraph: "It must be pointed out at the outset that, as has been found in paragraphs 36 to 38 of the present judgment, processing of personal data, such as that at issue in the main proceedings, carried out by the operator of a search engine is liable to affect significantly the fundamental rights to privacy and to the

industry-wide effort, which currently still requires robust research and innovation efforts for its development. [9] While significant volumes of data are often required for effectively training Gen AI models, large-scale web scraping of personal data should only be considered if there are no reasonable alternatives.

Furthermore, as the ICO notes, legitimate interest is likely to be the most practicable legal basis for training Gen AI models on personal data scraped from publicly accessible sources. Other legal bases such as consent appear impracticable (e.g. for seeking and withdrawing consent) and may result in incomplete or non-representative training data.

The risk-assessment requirement (or "balancing test", see next question) inherent in the legitimate interest basis also provides for more accountability and controls than any other basis.

2) **As we explain in the consultation, the legitimate interests test could be met if technical and organisational measures to limit the use of the Gen AI model are in place. Do you agree with the analysis we have presented? If yes, what measures should a developer implement to limit the ways in which its customers can use the Gen AI model?** *Yes.*

The final part of the 'three-part' test to meet the legitimate interest basis requires a balancing test as to whether individuals' interests override the interest being pursued, which effectively entails performing a risk assessment on the proposed processing activity. CIPL notes that in addition to the right to privacy, individuals have other fundamental rights that could be put at risk from inaccurate or biased outputs of AI models. It is critical that controllers have in place demonstrable safeguards to protect all these fundamental rights and mitigate risks when training Gen AI models using web-scraped personal data (e.g., performing risk assessments and DPIAs, ensuring data integrity, and providing appropriate transparency).

Regardless, 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, contract and other laws.

In many contexts, it may be advisable that developers of Gen AI models generally build in controls to prevent users from building detailed profiles of individuals, retrieving sensitive information about them, or generating their likenesses, as is already the case for a number of Gen AI tools in broad public use.

Model developers should also be incentivised to use privacy-enhancing technologies where feasible and appropriate, including, but not limited to, synthetic data and federated learning.

---

protection of personal data when the search by means of that engine is carried out on the basis of an individual's name, since that processing enables any internet user to obtain through the list of results a structured overview of the information relating to that individual that can be found on the internet — information which potentially concerns a vast number of aspects of his private life and which, without the search engine, could not have been interconnected or could have been only with great difficulty — and thereby to establish a more or less detailed profile of him. Furthermore, the effect of the interference with those rights of the data subject is heightened on account of the important role played by the internet and search engines in modern society, which render the information contained in such a list of results ubiquitous."

[9] For example, Google announced the first Machine Unlearning Challenge with a broad group of academic and industrial researchers aiming to advance the state of the art in machine unlearning and encourage the development of efficient, effective and ethical unlearning algorithms (https://blog.research.google/2023/06/announcing-first-machine-unlearning.html).

CIPL also asks the ICO to appropriately define key terms such as 'developer' and 'deployer'. The consultation refers to 'customers' and 'third parties', which we assume intend to refer to deployers, but may lead to confusion.

3) This is the first in a series of publications on the ICO's analysis of personal data processing involved in Gen AI. What aspects of this topic would you like us to consider in future publications?

- Address the broad range of tensions between AI technologies and data protection principles, such as the data minimisation principle, legal basis for sensitive data processing, transparency and explainability, rights of individuals (access, objection, deletion), rules on automated decision taking, and international data transfers.[10] In particular, analysis of how to address special category data when training Gen AI models, including whether there is a valid legal basis organisations can rely upon, given that article 9 of the UK GDPR prevents web scraping of special category data. Including guidance on how to address countries where images are considered sensitive data.
- Practical guidance on data ethics and the risk assessment/balancing test for Gen AI, that in addition to the risks to individuals takes also into account the risk of not deploying AI (loss of opportunity) and benefits from deploying AI.
- The role of certifications, codes of conduct, and other accountability tools for responsible Gen AI development and deployment.
- Best practices in red-teaming and other adversarial testing approaches for Gen AI, as well as other approaches to mitigate potential risks associated with potential use of sensitive training data.
- Possible unintended consequences arising from training Gen AI on web-scraped data, such as any limitations associated with this approach related to accuracy and bias.
- Specific considerations in the context of open source versus closed source Gen AI models.
- We recommend that the ICO focus on applications built upon Gen AI models and their uses, as well as the underlying model.
- The ICO could leverage and inform the work of the UK AI Safety Institute, and continue to cooperate with other regulators in this space through the Digital Regulation Cooperation Forum.

[10] See discussion of these issues in CIPL's *Artificial Intelligence and Data Protection: Delivering Sustainable AI Accountability in Practice, First Report* (https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_first_ai_report_-_ai_and_data_protection_in_tension__2_.pdf) and *Second Report* (https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_hard_issues_practical_solutions_english_feb20.pdf) .

Response by the Centre for Information Policy Leadership to the Information Commissioner's Office's Second Consultation on Purpose Limitation in the Generative AI Lifecycle

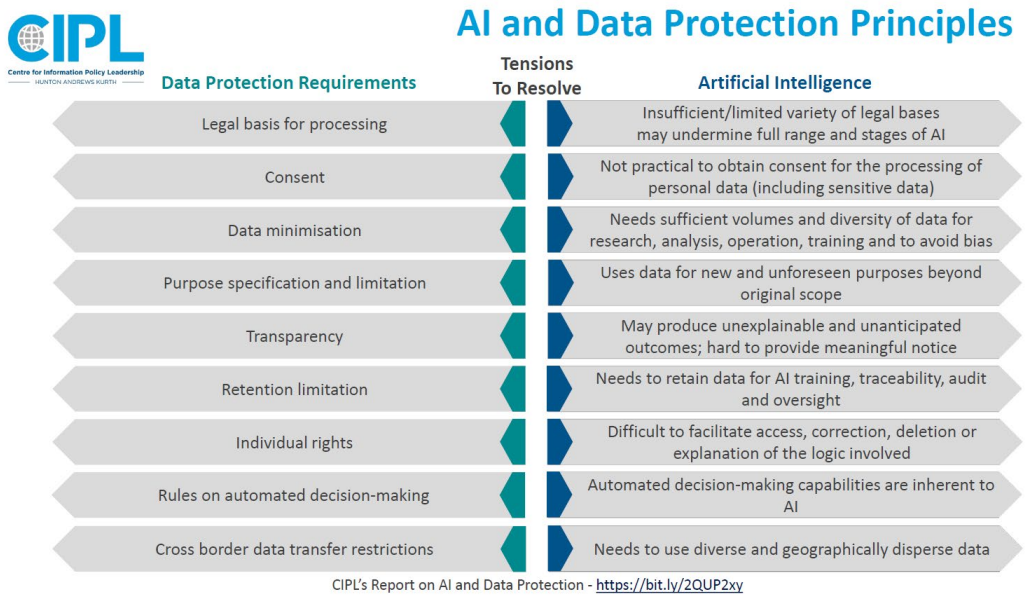
Submitted April 12, 2024

The Centre for Information Policy Leadership (CIPL) welcomes the opportunity to respond to the Information Commissioner's Office's (ICO) Consultation on how the data protection principle of purpose limitation should be applied at different stages in the generative AI lifecycle. 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.

CIPL particularly welcomes the ICO's efforts to clarify and evolve the interpretation of data protection principles for generative AI through this series of consultations. We encourage the ICO to continue to explore other areas, such as data minimisation, transparency, and data subject rights. CIPL has identified these tensions between data protection principles and AI in our work (please see the slide below), and we are pleased to see regulators responding to this. This is especially important in the UK where the Government's AI policy requires existing regulators to examine the application of their sectoral law to AI and produce guidance but we invite similar initiatives in other jurisdictions towards a balanced approach to AI and privacy.

[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.



CIPL's Report on AI and Data Protection - https://bit.ly/2QUP2xy

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

- CIPL agrees that careful consideration must be given to the purpose limitation principle when developing generative AI models or applications based on such models. The purpose limitation principle was introduced into data protection law to prevent a "free-for-all" in organisations' use and re-use of individuals' personal data. This objective remains important, but the rise of AI technologies requiring extensive amounts of data for training, development, and operation necessitates a closer look at the interpretation of this principle.
- We encourage the ICO to ensure that model developers have the ability to articulate purposes that are sufficiently broad and flexible for the range of potential applications for which they may be used.
- Furthermore, it is important to note that the principles of purpose specification and use limitation are not absolute. For example, the purpose limitation principle of the GDPR requires that personal data be "collected for specified, explicit and legitimate purposes, *and not further processed in a manner that is incompatible with those purposes."*[3] The *OECD Privacy Guidelines*, which underpin most modern data protection laws, contain similar language.[4] These principles are designed to limit unforeseen or hidden processing of data, and allow for compatible processing that serves the spirit of the principle while also enabling some flexibility. Ultimately, further processing based on "compatibility" should be allowed for future uses that are consistent with, can co-exist with, and do not undermine or negate the original purpose

[3] GDPR, "Art. 5 Principles relating to processing of personal data", May 2018, https://gdpr-info.eu/art-5-gdpr/

[4] OECD, "Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data", September 1980, https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188

- However, the purpose limitation principle alone does not provide comprehensive, appropriate protections against the potential risk of harms resulting from AI development and deployment. It must be backed by strong, accountability-based safeguards, including reasonable transparency and benefit/risk assessments that enable tailored mitigations to ensure that new uses do not expose the individual to unwarranted increased risks or adverse impacts.

2. We explain in this consultation that the purposes of generative AI model development and application development should be considered to be separate purposes. Do you agree with the analysis we have presented on this?

- Using data to train and develop a generative AI model must be seen as a separate purpose from using the data to develop and deploy a specific application. Model developers of GenAI models should assess and set out the purpose of each stage of the development and training of generative AI models and establish what personal data is needed for that purpose and ensure they have an appropriate legal basis to process the data. Application developers may proceed under separate and distinct processes and purposes that model developers may not have full insight into or control over. For this reason, application developers are the appropriate party to specify the purposes of processing personal data in connection with the application development.
- Furthermore, the initial training of the model cannot be seen as a singular, unrepeating stage of the AI development lifecycle; training is an iterative process and continues throughout the use of the AI model. Therefore, model developers may need to collect, retain, and use data beyond the initial training stage. Such ongoing use of data may be necessary to protect against bias and preserve the robustness, accuracy, and security of the model. It may also be necessary to use the learnings from the application development process to feed back into the model development processing to correct learned biases, as the ICO's flowchart recognises.
- While generative AI model development and application development are indeed distinct and sequential processes, in many instances the two stages may be integrated from the outset: for example, a model developer may build the model while simultaneously initiating work on potential applications. There should be an ability in such a case to use data across these processes.

3. Where the organisation developing a model is separate to that developing the application based on it: How can the model developer meaningfully communicate to the other organisation what personal data was used for the model training and why?

- Transparency is a core principle of accountability and is key to providing meaningful, user-centric communication regarding the use of personal data. As stated in CIPL's recent publication, *Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework*[5], many organisations developing AI models have already been

[5] CIPL, "Building Accountable AI Programs: Mapping Emerging Best Practices to the CIPL Accountability Framework", February 2024,

publishing transparent explanatory documents (e.g., model or system cards, technical reports) alongside product launches. These documents provide useful information regarding an AI model, such as how it was built, how it works, a summary of the types of data it was trained on, its intended use cases and contexts, key limitations, and basic performance metrics. Where possible and appropriate, such documents would also describe categories of personal data used in model training, including metadata on its key characteristics (e.g., what types of data are included in the dataset, where and how the data was collected, and which demographic groups are represented within it).

- At the same time, many generative AI models are trained on large and unstructured datasets. Any requirements to disclose details about the personal data contained therein could require indexing and other measures that might be in tension with data minimization principles.
- Model developers should take care not to disclose the specific data used to safeguard the privacy of individuals to whom the data pertains. Regulators should recognize the importance of safeguarding this information. Ultimately, transparency should be contextually appropriate, while also fulfilling transparency requirements under applicable laws and regulations. Thus for example, general purpose model developers should disclose information regarding how risks to data subjects were minimized in the context of model training.

4. Do you think the purpose of developing generative AI models requires the processing of personal data?

- In some circumstances, yes. Models may vary in the extent to which they rely on personal data, and many models process significant amounts of non-personal data, including agricultural and farming data, environmental data, chemical compounds, geographical and geological data, flight and shipping data, and more. At the same time, some models may require personal data to perform critical functions, such as reducing the risk of biased outputs. Furthermore, LLMs require data about people to learn how language incorporates concepts about relationships between people and the world. For instance, a model that generates text about a historical or current event will need to be able to correctly identify and use the proper names of people, places, and organisations involved in the event. Excluding, masking, or filtering out personal data from training data could severely hinder an LLM's ability to understand language and can impact the quality of the model. Furthermore, identifying personal data within a large dataset faces significant challenges, such as distinguishing fact from fiction, whether a person is living or dead, whether a word is a name, and what data is reasonably linked to any word that represents a living, non-public individual.
- At the same time, organisations should aim to limit the processing of personal data where feasible. For example, when practicable, organisations should consider employing anonymised or synthetic data enabled by privacy-enhancing technologies (PETs). In our report, *Privacy-Enhancing and Privacy-Preserving Technologies: Understanding the Role of*

https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_building_accountable_ai_programs_23_feb_2024.pdf.

PETs and PPTs in the Digital Age[6], we explored different privacy-enhancing technologies, and how they support data protection principles and legal compliance, as well as innovation. Several privacy-enhancing technologies can be used to reduce the necessity and resulting risks of processing personal data for developing generative AI models. For instance, prior to the development stage, safeguards may be put in place, such as filters or pattern recognition algorithms to reduce the amount of personal data in any downstream output; synthetic data that closely resembles real data may be used in some instances to train or validate the model without exposing sensitive information; in certain circumstances, differential privacy may be used to add noise during training to prevent identification of any individuals' data involved; and homomorphic encryption enables model training on encrypted data, keeping data secure throughout the training process. As stated in the ICO's 2023 Guidance on PETs, "effective anonymization" is vital, whereby regulators recognise that the risk of reidentification need not be reduced to zero in order to anonymise effectively.

5. **How can organisations who use personal data to train unspecified kinds of generative AI models comply with the purpose limitation principle?**
   - Training a general purpose, generative AI model is a purpose in and of itself, as developers are training the model to respond to different commands and generate a range of potential outputs. Because they may be building models with a range of potential future applications—some of which may be unknown at the time of model development,--the training and development of general purpose models *per se* should be recognised as a sufficiently specific, legitimate, and permissible purpose.
   - By their nature, base models often do not have a single, final purpose: developers who create both models and applications may not be able to demonstrate, or even identify, all possible and appropriate uses at the model development stage. However, transparency documents provided by developers, such as model cards or system cards, may serve as a purpose framework by indicating the range of purposes that the model should and should not be used for, to the extent possible. For example, the developer may outline a range of applications that the model is well suited for, given the type of data it was trained on. In the reverse, the developer may also be able indicate what purposes the model is not intended or suitable for.

6. **Do you consider the collection of training data and model development to fall under the same purpose?**
   - The collection of training data and model development should be considered different activities that are still pursuant to the same purpose where the collected data is being used to train, develop, and fine-tune the same model.

---

[6] CIPL, "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

7. **What aspects of generative AI development and deployment would need to be documented to make a purpose specific enough?**
    - CIPL agrees with the ICO's conclusion that model and application developers should be able to describe in plain language the stages of their model development process, and the purpose of collection and processing of personal data at each stage in relation to the model. Similarly, application deployers should provide clear explanation of how and why personal data is used to operate their applications.
    - Developers, deployers, and regulators alike must recognise that in the context of model development and deployment, organisations may need to use greater volumes of data than in other data processing contexts, while still satisfying the principles of purpose limitation and data minimisation. For example, a generative AI model that prohibits use by minors may need to be trained on age-related data to help identify and prevent its use by minors. Similarly, model developers may generally want their model to exclude certain types of data that do not fit within the model's intended purpose (e.g., children's, health, etc.), but to be able to identify and cleanse these data types from the model, developers must collect some of that data so that the model is sufficiently trained to exclude it, and may need to continue to collect and analyse such data over time to prevent "drift" toward collection of such data. This example demonstrates how narrowly-construed compliance with existing data protection laws (e.g., minimizing data collection as much as possible, limiting the further use of the data unless there is a "compatible" purpose with the initial processing, or deleting the data as soon as it is no longer necessary for the initial purpose) may be in tension with responsible AI development and deployment. The ICO can help developers and deployers navigate these tensions by issuing guidance that recognises and accounts for these tensions.

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