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

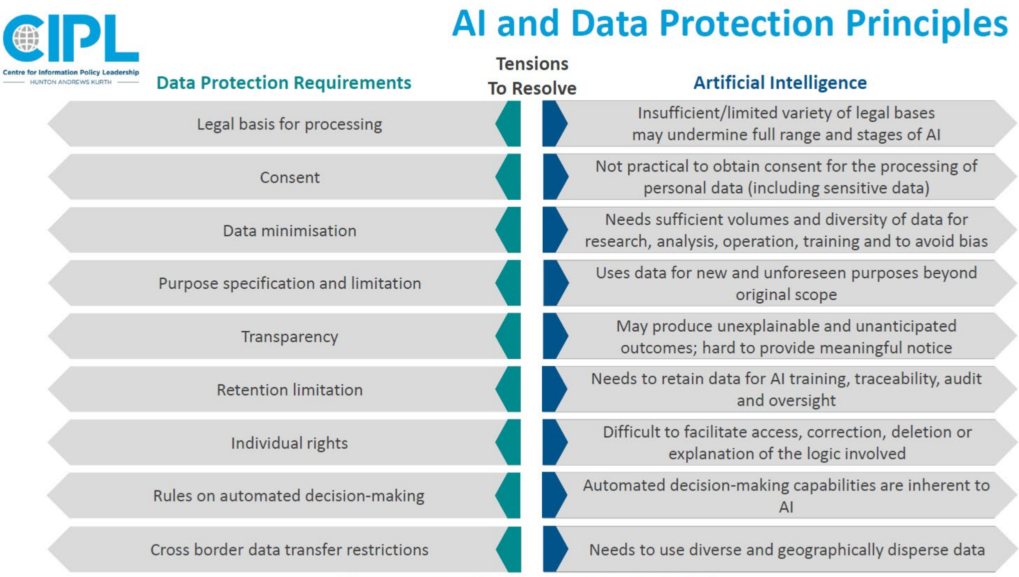
Submitted April 12, 2024

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
Centre for Information Policy Leadership
HUNTON ANDREWS KURTH

AI and Data Protection Principles

Data Protection Requirements	Tensions To Resolve	Artificial Intelligence
Legal basis for processing		Insufficient/limited variety of legal bases may undermine full range and stages of AI
Consent		Not practical to obtain consent for the processing of personal data (including sensitive data)
Data minimisation		Needs sufficient volumes and diversity of data for research, analysis, operation, training and to avoid bias
Purpose specification and limitation		Uses data for new and unforeseen purposes beyond original scope
Transparency		May produce unexplainable and unanticipated outcomes; hard to provide meaningful notice
Retention limitation		Needs to retain data for AI training, traceability, audit and oversight
Individual rights		Difficult to facilitate access, correction, deletion or explanation of the logic involved
Rules on automated decision-making		Automated decision-making capabilities are inherent to AI
Cross border data transfer restrictions		Needs to use diverse and geographically disperse data

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.