47th Closed Session of the Global Privacy Assembly

September 2025

Resolution on the collection, use and disclosure of personal data to pre-train, train and fine-tune AI models

This Resolution is submitted by the Office of the Australian Information Commissioner.

SPONSORS:

- Office of the Australian Information Commissioner

CO-SPONSORS:

- Garante per la Protezione dei Dati Personali (ITDPA, Italy)
- Office of the Privacy Commissioner New Zealand (OPC New Zealand)
- Office of the Privacy Commissioner for Bermuda (PrivCom Bermuda)
- Office of the Privacy Commissioner Canada (OPC Canada)
- Personal Data Protection Service of Georgia (PDPS Georgia)_
- Information and Privacy Commissioner of Ontario (IPC Ontario)
- Federal Data Protection and Information Commissioner of Switzerland (FDPIC Switzerland)
- National Privacy Commission of the Philippines (NPC Philippines)
- Croatian Personal Data Protection Agency (AZOP Croatia)
- Office of the Privacy Commissioner for Personal Data, Hong Kong, China (PCPD Hong Kong)
- European Data Protection Supervisor (EDPS Europa)
- Autorité de protection des données – Gegevensbeschermingsautoriteit (Belgium)
- Commission Nationale de l'Informatique et des Libertés (CNIL France)
- Commission for Personal Data Protection of the Republic of Bulgaria
- Data Protection Commission of Ireland

The 47th Global Privacy Assembly 2025:

ACKNOWLEDGING the various concerns expressed in the public debate regarding the ethical and legal implications of AI technologies, and particularly generative AI technologies, including their impact on fundamental rights and freedoms, notably in relation to the fundamental right to privacy and protection of personal data;

RECOGNISING that the development of AI technologies, and particularly generative AI technologies, often involves the large scale collection, use and disclosure of personal data, including special categories of data which have a strengthened protection under many legal frameworks, in order to pre-train, train and fine-tune AI models;

RECALLING the fundamental right to privacy and the protection of personal data implies a high degree of individual choice, control and agency over the collection, use and disclosure of personal data;

CONCERNED that the indiscriminate collection, use and disclosure of personal data to pre-train, train and fine-tune AI models without an appropriate legal basis is imperilling the fundamental right to privacy and the protection of personal data;

CONCERNED in particular about the indiscriminate and large scale collection of publicly available personal data without an appropriate legal basis, as defined under applicable data protection frameworks, thereby undermining the principle of fair and lawful processing and data subject expectations;

CONCERNED additionally about the use of personal data for secondary purposes related to the pre-training, training and fine-tuning of AI models without an appropriate legal basis;

NOTING with concern the emerging practice of distilling data, including personal data, from AI models without an appropriate legal basis;

RECALLING that data protection and privacy principles and current laws, including data protection and privacy laws, bills, statutes and regulations, apply to AI technologies, and particularly generative AI technologies, even as different jurisdictions continue to develop AI-specific laws and policies;

EMPHASISING the public interest in ensuring that AI technologies, particularly generative AI technologies, are developed and deployed in a way consistent with human rights standards, including data protection and privacy principles and laws, in order to ensure their trustworthiness and facilitate their adoption;

EMPHASISING the necessity of establishing comprehensive technical and organisational measures for data protection that encompass the entire lifecycle of the AI models, from initial data sourcing and all phases of model development (pre-train, train and fine-tuning) to their deployment and continuous monitoring, with particular attention to the enduring impact of the training data.

STRESSING that economic or political imperatives to accelerate AI technologies' development and adoption must maintain full respect for data protection and privacy principles and laws;

HIGHLIGHTING that personal data should only be used to pre-train, train or fine-tune AI models where it has been lawfully obtained and the collection, use and disclosure is consistent with the following non-exhaustive list of data protection principles:

1. Lawful and fair basis for processing

The collection of personal data for the pre-training, training and fine-tuning of AI models must be fair, lawful and transparent, regardless of whether that data is publicly accessible. The public availability of such data does not automatically imply a lawful basis for its processing, which must always be assessed in the light of the data subject's reasonable expectations of privacy.

In particular, the covert collection of personal data without the individual's knowledge could be considered unfair, except in limited and clearly justified circumstances where such collection is lawful, necessary and proportionate. A fair means of collection is unlikely to involve intimidation or deception.

It is also important to ensure that processing and outcomes are fair. Datasets should therefore be accurate, complete and representative of the purpose of the processing. Datasets should also be carefully assessed to avoid the inadvertent incorporation of any underlying unwanted bias.

2. Purpose specification and use limitation

Personal data must be used or disclosed solely for the purpose(s) for which it was originally collected. Further use or disclosure for a new purpose must be compatible with the initial purpose(s), or authorized by the data subject (consent), or authorised by law. Further use or disclosure should be legitimate, consistent with the reasonable expectations of the individual, and have a lawful basis.

3. Data minimisation

The collection and use of personal data must be limited to only what is reasonably necessary to fulfil a legitimate, specific and explicit purpose.

In assessing the necessity of the collection and use of personal data, consideration should be given to whether the AI model can be trained without the collection or use of personal data, or whether technical or organisational measures could be put in place to minimise the amount of personal data collected or used (e.g., anonymisation, pseudonymisation and privacy enhancing technologies).

4. Transparency

The use of personal data to pre-train, train or fine-tune AI models must be transparent. AI developers and deployers must implement adequate notice and transparency measures, including by providing in a timely manner clear, concise, intelligible and easily accessible information about the collection and use of personal data, how it will be held, when it can be disclosed, and the purposes for which it is collected. This should also include clear, accessible and meaningful information about the categories of data processed, data sources, retention periods and third-party recipients.

5. Accuracy

Developers must take reasonable steps to ensure accuracy in pre-training, training or fine-tuning AI models, commensurate with the likely increased level of risk in an AI context, including through using high quality datasets and undertaking appropriate testing to ensure a high degree of accuracy in the model's outputs. The use of disclaimers to signal where AI models may require careful consideration and additional safeguards for certain high privacy risk uses may be appropriate.

6. Data security:

Entities must implement appropriate technical and organizational measures to protect personal data used in AI model training at every stage of model development, including pre-training, training, and fine-tuning. Personal data should be secured against unauthorized access, breaches, and unintended disclosure throughout the AI development lifecycle. In particular, effective safeguards should be in place to prevent and detect attempts to extract or reconstruct personal data from trained AI models.

Entities involved in the collection, use, or disclosure of personal data for AI model development are encouraged to adopt privacy-enhancing tools and technologies (PETs), including pseudonymisation and anonymisation techniques. PETs should be used to minimise data exposure, reduce re-identification risks and support compliance with data protection principles throughout the AI lifecycle.

7. Accountability and privacy by design:

Those who collect or use personal data for AI training should be accountable for complying with these principles. AI developers and AI deployers should embed data protection by design and by default into AI systems, for example, by planning for minimization, transparency, and security from the outset. They should document and be able to demonstrate the steps taken to ensure compliance (such as conducting data protection impact assessments or similar risk assessments for high-risk AI projects, and implementing governance measures to oversee AI data practices).

8. Rights of data subjects

Where provided by law, developers and deployers of AI models must ensure that individuals, as data subjects, are able to access information about the collection, use and disclosure of their personal data, particularly where it has been used in the pre-training, training and fine-tuning of an AI model, and where that data has been obtained from a variety of sources. Individuals should be permitted to request deletion of that data, object to use of their data, and otherwise exercise their data protection rights. AI developers and deployers shall implement the appropriate technical and organisational measures to ensure that affected individuals are able to exercise these rights where provided by law. Consideration should be had regarding the individual's capacity and/or vulnerability and how this impacts their ability to provide informed consent in relation to the collection, use and disclosure of their personal data.

HIGHLIGHTING the constrained resources of privacy and data protection authorities compared to the exceptionally fast technological developments taking place at a global level by well-resourced entities;

REAFFIRMING the 41st Global Privacy Assembly's *International Resolution on Privacy as a Fundamental Human Right and Precondition For Exercising Other Fundamental Rights* that to build trust in our digital society, accelerate innovation, and protect human dignity, generative AI should be human centric, based on democratic values, and should recognize privacy as a fundamental human right, vital to the protection of other rights and freedoms;

RECOGNISING that AI has the potential to benefit society, by improving efficiency and productivity for the community, when deployed safely and in a rights-preserving way.

RECALLING the 45th Global Privacy Assembly's *Resolution on Generative Artificial Intelligence Systems* commitment to ensuring the application and enforcement of data protection and privacy legislation in the context of generative AI technologies, including the applicable principles and rights set out in this resolution;

RECALLING the 45th Global Privacy Assembly's Resolution on AI and Employment, with regard to training AI whereby committing to reduce and mitigate biases or discrimination, both direct and indirect, when developing and deploying an AI system in the employment context, including by taking reasonable steps to ensure personal data used in the training of a system and in solely automated decision making is representative to the context in which the system will be used, accurate and regularly updated, and implementing appropriate technical and organisational measures to ensure, in particular, that factors in recruitment and work management systems which result in inaccuracies in personal data are corrected and the risk of errors is minimised, as well as compliance with the applicable domestic laws;

RECALLING the need to recognise that DPAs are at the forefront of shaping data governance to address the evolving challenges of AI, also in cooperation with other authorities, as underlined in the *Statement on the Role of Data Protection Authorities in Fostering Trustworthy AI*, adopted by the G7 Data Protection and Privacy Authorities Roundtable in Rome on 11 October 2024;

BEING AWARE of the relevance of a common legal framework based on shared values and principles, such as the Framework Convention of the Council of Europe on Artificial Intelligence, opened for signature on September 5, 2024 which, together with Convention 108+ for the protection of individuals with regard to the processing of personal data represent complementary tools for the protection of fundamental rights at global level, being both open to the ratification of all countries;

The 47th Global Privacy Assembly therefore resolves to:

1. Agree and emphasise that the collection, use and disclosure of personal data to pre-train, train and fine tune AI models is within the scope of existing and emerging data protection and privacy principles and laws, including the data protection principles highlighted in this resolution, and all entities who collect, use and disclose personal data to pre-train, train and fine tune AI models must adhere to them;

2. Promote these principles and engage with policy makers and international organisations such as the OECD, the Council of Europe and the UN and other relevant stakeholders to

raise awareness, educate and supervise AI developers and deployers in their GPA member jurisdictions and globally to ensure that the collection, use and disclosure of personal data to pre-train, train and fine-tune AI models is compatible with data protection and privacy principles and laws;

3. Coordinate enforcement efforts on generative AI technologies and in particular to ensure a consistent standard of data protection and privacy is applied to the pre-training, training and fine-tuning of AI models and encourage cross-border cooperation among DPAs for more effective oversight of global AI services;

4. Commit to sharing ongoing developments on education, compliance and enforcement efforts on generative AI technologies, with a view to fostering coherence of regulatory approaches and enhancing mutual assistance.