JANUARY 2026

Principles for the Responsible Use of Artificial Intelligence

Information and Privacy Commissioner of Ontario
Commissaire à l'information et à la protection de la vie privée de l'Ontario

Ontario

Ontario Human Rights Commission
Commission ontarienne des droits de la personne

Contents

Preface

Artificial intelligence (AI) has the potential to significantly enhance the lives of all Ontarians. To realize this potential, it is imperative that AI systems are developed, acquired, used, and decommissioned in a manner that upholds safeguards for human rights, including the right to privacy. Accordingly, the Office of the Information and Privacy Commissioner of Ontario (IPC) and the Ontario Human Rights Commission (OHRC) continue to emphasize the importance of responsible and trustworthy AI adoption by the Ontario public sector and the broader public sector. The principles outlined herein, jointly developed by the IPC and OHRC, identify key concepts that will ground our assessment of organizations' adoption of AI systems consistent with privacy and human rights obligations.

The IPC–OHRC Principles represent a versatile and scalable foundation for responsible AI governance. These principles assist institutions in responsibly implementing AI innovations while ensuring the protection of privacy, human rights, human dignity, and public trust for Ontarians.

Institutions are strongly encouraged to adopt the IPC–OHRC AI Principles to ensure that their use of AI systems is responsible, transparent, and compliant with Ontario's human rights and privacy laws. These principles offer a clear, credible, and robust framework for assessing risk, guiding system design and deployment, and embedding accountability throughout the AI life cycle. By adhering to the IPC-OHRC principles, institutions can effectively safeguard individuals and communities from potential harms, show their commitment to fairness and substantive equality, and improve public trust. Ultimately, implementing the IPC–OHRC AI Principles helps ensure that AI systems uphold the rights and dignity of people affected, while fostering responsible innovation throughout the development, provision, and use of AI systems.

Organizations in Canada and internationally are increasingly implementing AI principles to address the challenges associated with adopting AI systems. Notable initiatives include the European Union (EU) Ethics Guidelines for Trustworthy AI,[1] the Group of Seven (G7) Hiroshima Process establishing International Guiding Principles for Organizations Developing Advanced AI Systems,[2] and the Organization for Economic Cooperation and Development (OECD) AI Principles.[3] In Canada, the federal government has introduced an AI strategy for the federal public service,[4] and Ontario has established a directive for all provincial ministries and agencies regarding the responsible use of AI.[5] The IPC–OHRC principles presented in this document are designed to complement these provincial, national, and international principles, while emphasizing the protection of human rights, including privacy laws.

1 Ethics guidelines for trustworthy AI: **https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai**

2 Hiroshima Process. International Guiding Principles for Organizations Developing Advanced AI systems: **https://www.soumu.go.jp/hiroshimaaiprocess/pdf/document04_en.pdf**

3 OECD AI Principles: **https://www.oecd.org/en/topics/ai-principles.html**

4 AI Strategy for the Federal Public Service 2025-2027: **https://www.canada.ca/en/government/system/digital-government/digital-government-innovations/responsible-use-ai/gc-ai-strategy-overview.html#toc-3**

5 Ontario Responsible Use of Artificial Intelligence Directive: **https://www.ontario.ca/page/responsible-use-artificial-intelligence-directive**

What is artificial intelligence?

Ontario's *Enhancing Digital Security and Trust Act* (EDSTA), defines an "artificial intelligence system" as:

a) a machine-based system that, for explicit or implicit objectives, infers from the input it receives in order to generate outputs such as predictions, content, recommendations or decisions that can influence physical or virtual environments, and

b) such other systems as may be prescribed.

For the purposes of the joint IPC–OHRC principles, we have adopted this EDSTA definition of AI systems. This definition is consistent with the OECD definition of an AI system.[6] The OECD's definition was created following a global consensus-seeking process with an extensive range of interested parties, and as such, represents a broad conception of what an AI system might include.

For clarity, the OECD definition applies to, among other things:

- automated decision-making systems
- systems which are designed to undertake activities that are typically performed using human intelligence and skills
- generative AI systems
- foundational large language models (LLMs) as well as their applications
- traditional AI technologies (e.g., spam filters or other cyber security resources, computer vision systems)
- any other emerging innovative uses of AI technologies

The life cycle of AI

The life cycle of AI includes the following stages:

1. **Design, data, and modelling**: This is the first stage in which system objectives, underlying assumptions, context, and requirements are specified. Data to power the AI system is collected, processed, and checked for quality. The AI system developers then create or select a model or algorithm that is trained or calibrated against the data set.
2. **Verification and validation**: At this second stage, developers assess their model for its performance against objectives. This could include assessing false positives, false negatives, and/or performance under a variety of conditions.
3. **Deployment**: The model and its overall system are launched for use in an environment. The system may begin to monitor the environment, assess collected data using its models, and generate outputs such as predictions, categorizations, decisions, and assessments.

6 OECD Definition: **https://oecd.ai/en/wonk/ai-system-definition-update**

4. **Operation and monitoring**: The AI system is in operation, with its outputs being used in service of the AI system's objectives. The system is monitored based on performance and quality evaluation criteria. Based on monitoring results, the system operators may take the system back to earlier phases to re-evaluate its design and training.[7]

5. **Decommissioning**: The life cycle ultimately extends to the decommissioning of an AI system. Decommissioning may take place because an AI system has reached its end of life or because it is routinely exhibiting unexpected outputs, and its behaviour cannot be corrected. The AI system, and the data used, including previously produced outputs, are retained as lawfully required to justify, rationalize, or explain past actions, as well as to assess the unexpected outputs and how individuals or communities have potentially been affected by them.

Each stage of the AI's life cycle should be assessed against the relevant principles in this document. Assessments at relevant stages should be conducted pursuant to an institution's role as a developer, provider,[8] or user[9] of a given AI system.

Principles for responsible use of AI

These principles are to be considered interconnected and of equal importance.

Principle: Valid and reliable

> AI systems must exhibit valid, reliable, and accurate outputs for the purpose(s) for which they are designed, used, or implemented.

To be valid, AI systems must meet independent testing standards and be shown, using objective evidence, to fulfil the intended requirements for a specified use or application. They must be proven to be reliable by performing consistently, as required, over a specified duration, and in the environments in which they are intended to be used. They must also be robust enough to maintain that level of performance across various other operating conditions, particularly in situations in which experiences and outcomes may differ for Ontario's diverse communities.

Validity and reliability standards contribute to the accuracy of observations, computations, or estimates so that results can be reasonably accepted as being true. However, the accuracy of results also depends heavily on the accuracy, completeness, and quality of the input data provided to the AI system. Even a highly valid and reliable tool can yield poor outcomes if it is provided with inaccurate, biased, or incomplete data.

---

7 Organization for Economic Co-Operation and Development. "The Technical Landscape." Artificial Intelligence in Society. June 11, 2019. **https://www.oecd.org/en/publications/artificial-intelligence-in-society_eedfee77-en.html**

8 A provider is defined as individuals or organizations that develop (including training) AI systems, or that put such services onto the market.

9 A user is defined as a staff member or agent of an organization who makes use of an AI system in the course of their institutional activities. Users do not design or provide the system, but they interact with, rely on, or apply its outputs to support decision-making, deliver services, or carry out organizational functions.

An AI system, therefore, should pass validity and reliability assessments prior to being deployed and be regularly assessed throughout its life cycle to confirm that it continues to produce accurate results and to operate as expected in a variety of circumstances.

Principle: Safe

> AI must be developed, acquired, adopted, and governed to prevent harm or unintended harmful outcomes that infringe upon human rights, including the right to privacy and non-discrimination.

AI systems should be monitored to support, among other considerations, human life, physical and mental health, economic security, and the environment. AI systems should be monitored and evaluated throughout their life span to confirm that they can withstand unexpected events or deliberate efforts that cause harm. This will, in part, require demonstrating that the AI systems have robust cyber security protection, and that human rights and privacy safeguards are firmly in place.

Any new use of a given AI system should undergo a comprehensive assessment process to ensure it will constitute a safe use in the new context. Safe AI systems must also make evident when they are producing unexpected outputs. AI systems should be temporarily or permanently turned off or decommissioned when they become unsafe, and any negative impacts to individuals and groups must be reviewed accordingly.

Principle: Privacy protective

> AI should be developed using a privacy by design approach. Developers, providers, or users of AI systems should take proactive measures to protect the privacy and security of personal information and support the right of access to information from the very outset.

AI systems should be developed using a privacy by design approach that anticipates and mitigates privacy risks to individuals and groups. This approach ensures that privacy protections are embedded into the system from the outset, proactively safeguard personal data, and respect the privacy of all individuals, especially those who are vulnerable or unable to provide informed consent. AI systems often interact with, or process, significant volumes of personal information in their development, training, or operation. The privacy protection principle requires clear lawful authority to collect, process, retain, and use these data. Accordingly, developers, providers, or users of AI systems must comply with applicable federal or provincial privacy laws, directives, regulations, or other legal instruments.[10]

10 AI systems can pose fundamental challenges to principles that have traditionally undergirded privacy legislation. The principle of limiting collection is challenged given that AI systems routinely require large and diverse volumes of data and information to best function. Data and information are sometimes re-used to train AI systems, placing pressure on the principle of purpose limitation, and what is learned during the training phases of AI systems may be retained after the training data is deleted with the effect of challenging the principle of limiting retention. Finally, even where organizations have attempted to anonymize information, the resulting data may sometimes be re-identified by AI systems.

Any use of personal information should be limited to what is required to fulfill the intended purpose. Institutions developing, providing, or using AI systems should reduce the need for large volumes of personal information using privacy enhancing technologies including de-identification methods or the use of synthetic data.

Privacy protective AI systems must build in measures to adjust the training data to mitigate any inherent bias and to ensure the accuracy of AI outputs, particularly where consequential decisions or inferences are being made about individuals or groups based on these outputs.

Individuals should be informed whether and when their personal information is being used in the development, refinement, or operation of an AI system, as well as the purpose and intended use of the AI system. Where appropriate, individuals should be provided with an opportunity to access or correct their personal information, including information about them generated by an AI system. Individuals should be provided with at least a right of review for automated decision processes that do not involve high risk, and the choice of opting out of high-risk automated decision processes that can materially impact an individual's well-being in preference of a human decision maker.[11]

AI systems must also be designed to protect the security of personal information from unauthorized access. Strong security safeguards are essential to ensure that personal information is protected from unauthorized access or misuse through the AI's life cycle.

Principle: Human rights affirming

> Human rights are inalienable, and protections must be built into the design of AI systems and procedures. Institutions using AI systems must prevent and remedy discrimination effectively and ensure that benefits from the use of AI are universal and free from discrimination.

Human rights law requires that developers, providers, and institutions ensure that they do not infringe substantive equality rights. This can be done by proactively identifying and addressing systemic discrimination in the design and deployment of AI systems on grounds protected under the *Ontario Human Rights Code (Code)*.[12] Institutions should take active measures to mitigate the discriminatory impacts present in AI systems and their associated data sets, such as adjusting training data to resolve any inherent biases detected through ongoing monitoring. In addition, institutions should avoid the uniform use of AI systems with diverse groups. Such a use, though seemingly neutral, may actually result in adverse impact discrimination.

11 Impact assessments are among the leading strategies to identify and assess for risk associated with AI systems. The OHRC (with the Law Commission of Ontario) and the IPC have impact assessments at their respective websites to identify, assess, and mitigate against human rights and privacy risks. For OHRC see Human Rights AI Impact Assessment: **https://www3.ohrc.on.ca/en/human-rights-ai-impact-assessment**. For IPC see Privacy Impact Assessment Guide: **https://www.ipc.on.ca/en/resources/planning-success-privacy-impact-assessment-guide-ontarios-public-institutions**.

12 Ontario Human Rights Code: **https://www.ontario.ca/laws/statute/90h19**.

Institutions have both privacy and human rights obligations to ensure that the collection, processing, and sharing of personal information or pseudonymous or anonymous data does not contribute to or reinforce existing inequalities or discrimination.

Likewise, government and governmental actors must comply with the rights guaranteed under the *Canadian Charter of Rights and Freedoms*, including the rights to freedom of expression, peaceful assembly, and association. This includes, but is not limited to, ensuring that AI systems do not unduly target participants in public or social movements, or subject marginalized communities to excessive surveillance that impedes their ability to freely associate with one another.

Principle: Transparent

> Institutions that develop, provide, and use AI must ensure that these AI systems are visible, understandable, traceable, and explainable to others.

Transparency involves providing clear notice about the use of AI systems, and adopting policies and practices that make visible, explainable, and understandable how AI systems work. Institutions developing, providing, or using AI must also ensure that AI systems are traceable and explainable. Transparency fosters public trust by enabling interested parties to understand how an AI system functions, how it produces its outputs, and the measures being taken to ensure that the AI system operates safely and accurately. Transparency consists of the following characteristics.

First, AI systems must be visible. This means that institutions should provide a public account that explains the operation of the system throughout its life cycle, from design and development to deployment and eventual decommissioning. This documentation may include privacy impact assessments, algorithmic impact assessments, or other relevant materials. Institutions must also be transparent about the sources of any personal data collected and used to train or operate the system, the intended purposes of the system, how it is being used, and the ways in which its outputs may affect individuals or communities. Importantly, this documentation should be written in clear, accessible language that avoids unnecessary jargon and technical complexity. Furthermore, institutions must notify individuals when they are interacting with an AI system and when any information presented to them has been generated by AI systems.

Second, AI systems must be understandable. This means that institutions must be able to explain how the technology operates and why errors may occur. To achieve this, they should document and retain sufficient technical information about the systems they are using so they can provide a full and transparent accounting of the basis on which decisions or actions were taken.

AI system's vendors should design and communicate about their AI systems in such a way that allows institutions that deploy and use them to understand how the AI system operates and how and why its outputs are generated as they are.

Third, AI systems must be explainable. This means institutions must be able to describe both the process (how) and the rationale (why) behind the outputs AI systems generate. This information should be communicated in a clear and accessible manner. The level of detail may

vary according to the audience — whether it is directed to the public, non-experts, individuals, or groups directly impacted by AI systems, or regulators overseeing institutional practices.

Fourth, AI systems must be traceable, meaning it must be possible for institutions to collect a thorough account of how the system operates, which can include:

- model details, such as the intended use of an AI system, type(s) of algorithm or neural network, hyperparameters, as well as pre- and post-processing steps
- training and validation data, including details on data gathering processes, data composition, acquisition protocols, and data labelling information
- AI tool monitoring details, which can include performance metrics, failures, and periodic evaluations[13]

Principle: Accountable

> Institutions should implement a robust internal governance structure with clearly defined roles, responsibilities, and oversight procedures, including a human-in-the-loop approach, to ensure accountability throughout the entire life cycle of their AI systems.

Incorporating robust internal governance structures, including a human-in-the-loop approach, ensures that human oversight is maintained throughout the life cycle of the AI system and allows for real-time intervention as needed.

Up front risk assessments should be carried out to identify and assess risks associated with the AI system, and to develop measures necessary to mitigate against them. Such assessments should include privacy and human rights impact assessments, algorithmic impact assessments, and others as relevant and appropriate.

Institutions should designate a person or persons responsible for overseeing the development, deployment, and/or use of an AI system, and for pausing or decommissioning an AI system that produces unsafe outputs or begins to operate in ways which are not valid or reliable.

Institutions should document their decisions about design and application choices in relation to AI systems. Where such a decision impacts specific groups or communities, they should be meaningfully informed and provided an opportunity to challenge that decision and any related outputs or results and seek recourse accordingly.

Institutions should be prepared to explain and provide plain language documentation on how the AI system works to an independent oversight body, upon request, and undertake any remedial or corrective actions as directed. Institutions must establish a mechanism to receive and respond to privacy, transparency, or human rights questions or concerns, as well as freedom of information requests, or to any challenges concerning how the AI system arrived at a decision or was used during a decision-making process.

13 European Parliamentary Research Service. 2023. “Artificial Intelligence in Healthcare: Applications, risks, and ethical and societal impacts.”

Members of institutions should be empowered through safe whistleblowing protections to report instances where an AI system does not comply with legal, technical, or policy requirements. Whistleblowers should be able to report non-compliance to an independent oversight body responsible for reviewing or overseeing the AI system, without fear of reprisal. Institutions should be subject to review by an independent oversight body with authority to enforce this and the other AI principles and require the organization to undertake remedial or corrective actions associated with the AI system.

Principles for the Responsible Use of Artificial Intelligence

Information and Privacy Commissioner of Ontario

Commissaire à l'information et à la protection de la vie privée de l'Ontario

2 Bloor Street East, Suite 1400
Toronto, Ontario
Canada M4W 1A8

Tel.: (416)-326-3333
Website: www.ipc.on.ca
Email: info@ipc.on.ca

January 2026

Ontario Human Rights Commission

Commission ontarienne des droits de la personne

Office of the Chief Commissioner

180 Dundas Street West, Suite 900
Toronto ON M7A 2G5

Tel.: (416) 314-4537
Fax: (416) 314-7752