# Voluntary Code of Conduct on the Responsible Development and Management of Advanced Generative AI Systems

**From: Innovation, Science and Economic Development Canada**

September 2023

Advanced AI systems capable of generating content — such as ChatGPT, DALL·E 2, and Midjourney — have captured the world's attention. The general-purpose capabilities of these advanced AI systems offer enormous potential for innovation in a number of fields, and they are already being adopted and put to use in a variety of contexts. These advanced systems may be used to perform many different kinds of tasks — such as writing emails, answering complex questions, generating realistic images or videos, or writing software code.

While they have many benefits, advanced generative AI systems also carry a distinctly broad risk profile, due to the broad scope of data on which they are trained, their wide range of potential uses, and the scale of their deployment. Systems that are made publicly available for a range of different uses can present risks to health and safety, can propagate bias, and carry the potential for broader societal impacts, particularly when used by malicious actors. For example, the capability to generate realistic images and video, or to impersonate the voices of real people, can enable deception at a scale that can damage important institutions, including democratic and criminal justice systems. These systems may also have important implications for individual privacy rights, as highlighted in the G7 Data Protection and Privacy Authorities' Statement on Generative AI.

Generative systems can also be adapted by organizations for specific uses – such as corporate knowledge management applications or customer ser -which generally present a narrower range of risks. Even so, there a of steps that need to be taken to ensure that risks are appropriately i nd mitigated.

To address and mitigate these risks, signatories to this code commit to adopting the identified measures. The code identifies measures that should be applied in advance of binding regulation pursuant to the *Artificial Intelligence and Data Act* by all firms developing [1] or managing the operations [2] of a generative AI system with general-purpose capabilities, as well as additional measures that should be taken by firms developing or managing the operations of these systems that are made widely available for use, and which are therefore subject to a wider range of potentially harmful or inappropriate use. Firms developing and managing the operations of these systems both have important and complementary roles. Developers and managers need to share relevant information to ensure that adverse impacts can be addressed by the appropriate firm.

While the framework outlined here is specific to advanced generative AI systems, many of the measures are broadly applicable to a range of high-impact AI systems and can be readily adapted by firms working across Canada's AI ecosystem. It is also important to note that this code does not in any way change existing legal obligations that firms may have – for example, under the *Personal Information Protection and Electronic Documents Act*.

In undertaking this voluntary commitment, developers and managers of advanced generative systems commit to working to achieve the following outcomes:

- **Accountability** – Firms understand their role with regard to the systems they develop or manage, put in place appropriate risk management systems, and share information with other firms as needed to avoid gaps.
- **Safety** – Systems are subject to risk assessments, and mitigations needed to ensure safe operation are put in place prior to deployment.
- **Fairness and Equity** – Potential impacts with regard to fairness and equity are assessed and addressed at different phases of developmen deployment of the systems.
- **Transparency** – Sufficient information is published to allow co make informed decisions and for experts to evaluate whether risks have

been adequately addressed.

- **Human Oversight and Monitoring** – System use is monitored after deployment, and updates are implemented as needed to address any risks that materialize.
- **Validity and Robustness** – Systems operate as intended, are secure against cyber attacks, and their behaviour in response to the range of tasks or situations to which they are likely to be exposed is understood.

Signatories also commit to support the ongoing development of a robust, responsible AI ecosystem in Canada. This includes contributing to the development and application of standards, sharing information and best practices with other members of the AI ecosystem, collaborating with researchers working to advance responsible AI, and collaborating with other actors, including governments, to support public awareness and education on AI. Signatories also commit to develop and deploy AI systems in a manner that will drive inclusive and sustainable growth in Canada, including by prioritizing human rights, accessibility and environmental sustainability, and to harness the potential of AI to address the most pressing global challenges of our time.

## Signatories to the code of conduct

**Filter items** □ Showing 1 to 10 of 14 entries**Show** 10 ⌄ **entries**

| **Signatories** ↑ ↓ |
|---|
| Ada |
| AlayaCare |
| Alberta Machine Intelligence Institute (Amii) |
| Appen |
| BlackBerry |
| Cohere |

| **Signatories** ↑ ↓ |
|---|
| Council of Canadian Innovators |
| Coveo |
| Mila |
| OpenText |

1 2 Next →

## Measures to be undertaken pursuant to the

# Code of Conduct

| | | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| **Principle** | **Measures** | **Developers** | **Managers** | **Developers** | **Managers** |
| Accountability | **Implement a comprehensive risk management framework proportionate to the nature and risk profile of activities.** This includes establishing policies, procedures, and training to ensure that staff are familiar with their duties and the organization's risk management practices. | Yes | Yes | Yes | Yes |

| Principle | Measures | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| | | **Developers** | **Managers** | **Developers** | **Managers** |
| | **Share information and best practices on risk management with firms playing complementary roles in the ecosystem.** | Yes | Yes | Yes | Yes |
| | **Employ multiple lines of defence**, including conducting third-party audits prior to release. | | | Yes | |

| Principle | Measures | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| | | Developers | Managers | Developers | Managers |
| Safety | **Perform a comprehensive assessment of reasonably foreseeable potential adverse impacts**, including risks associated with inappropriate or malicious use of the system. | Yes | Yes | Yes | Yes |
| | **Implement proportionate measures to mitigate risks of harm**, such as by creating safeguards against malicious use. | Yes | | Yes | |

| **Principle** | **Measures** | **All advanced generative systems** | | **Advanced generative systems available for public use** | |
|---|---|---|---|---|---|
| | | **Developers** | **Managers** | **Developers** | **Managers** |
| | **Make available to downstream developers and managers guidance on appropriate system usage,** including information on measures taken to address risks. | Yes | | Yes | |
| Fairness and Equity | **Assess and curate datasets used for training** to manage data quality and potential biases. | Yes | | Yes | |
| | **Implement diverse testing methods and measures to assess and mitigate risk of biased output prior to release**. | Yes | | Yes | |

| | | **All advanced generative systems** | | **Advanced generative systems available for public use** | |
|---|---|---|---|---|---|
| **Principle** | **Measures** | **Developers** | **Managers** | **Developers** | **Managers** |
| Transparency | **Publish information on capabilities and limitations of the system.** | | | Yes | |
| | **Develop and implement a reliable and freely available method to detect content generated** by the system, with a near-term focus on audio-visual content (e.g., watermarking). | | | Yes | |
| | **Publish a description of the types of training data used to develop the system**, as well as measures taken to identify and mitigate risks. | | | Yes | |

| Principle | Measures | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| | | Developers | Managers | Developers | Managers |
| | **Ensure that systems that could be mistaken for humans are clearly and prominently identified as AI systems.** | | Yes | | Yes |

| | | **All advanced generative systems** | | **Advanced generative systems available for public use** | |
|---|---|---|---|---|---|
| **Principle** | **Measures** | **Developers** | **Managers** | **Developers** | **Managers** |
| Human Oversight and Monitoring | **Monitor the operation of the system for harmful uses or impacts after it is made available**, **including through the use of third-party feedback channels,** and inform the developer and/or implement usage controls as needed to mitigate harm. | | Yes | | Yes |
| | **Maintain a database of reported incidents after deployment, and provide updates as needed to ensure effective mitigation measures**. | Yes | | Yes | |

| | | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| **Principle** | **Measures** | **Developers** | **Managers** | **Developers** | **Managers** |
| Validity and Robustness | **Use a wide variety of testing methods across a spectrum of tasks and contexts prior to deployment** to measure performance and ensure robustness. | Yes | | Yes | |
| | **Employ adversarial testing (i.e., red-teaming) to identify vulnerabilities.** | Yes | | Yes | |
| | **Perform an assessment of cyber-security risk and implement proportionate measures to mitigate risks**, including with regard to data poisoning. | Yes | | Yes | Yes |

| Principle | Measures | All advanced generative systems | | Advanced generative systems available for public use | |
|---|---|---|---|---|---|
| | | Developers | Managers | Developers | Managers |
| | **Perform benchmarking to measure the model's performance against recognized standards**. | Yes | | Yes | |

# Footnotes

[1] Development includes methodology selection, collection and processing of datasets, model building, and testing.

[2] Managing the operations includes putting a system into operation, controlling the parameters of its operation, controlling access, and monitoring its operation.

**Date modified:**
2023-10-20

**ED-ISDE**

**ontact ISED**