Government of Canada Gouvernement du Canada

Canada.ca › About government › Government in a digital age › Digital government innovation › Responsible use of artificial intelligence (AI)

Guide on the use of Generative AI

Table of contents

Overview

Generative artificial intelligence (AI) tools offer many potential benefits to Government of Canada (GC) institutions. Federal institutions should explore potential uses of generative AI tools for supporting and improving their operations. However, because these tools are evolving, they should not be used in all cases. Federal institutions must be cautious and evaluate the risks before they start using them. The use of these tools should be restricted to instances where risks can be effectively managed.

This document provides preliminary guidance to federal institutions on their use of generative AI tools. This includes instances where these tools are deployed by federal institutions. It provides an overview of generative AI, identifies challenges and concerns relating to its use, puts forward principles for using it responsibly, and offers policy considerations and best practices.

This guide also seeks to raise awareness and foster coordination among federal institutions. It highlights the importance of engaging key stakeholders before deploying generative AI tools for public use and before using them for purposes such as service delivery. These stakeholders include legal counsel, privacy and security experts, and the Office of the Chief Information Officer at the Treasury Board of Canada Secretariat (TBS).

The guide complements and supports compliance with many existing federal laws and policies, including in areas of privacy, security, intellectual property, and human rights. The guide is intended to be evergreen as TBS recognizes the need for iteration to keep pace with regulatory and technological change.

What is generative AI?

The *Directive on Automated Decision-Making* defines AI as information technology that performs tasks that would ordinarily require biological brainpower to accomplish, such as making sense of spoken language, learning behaviours, or solving problems.

Generative AI is a type of AI that produces content such as text, audio, code, videos and images. [1] This content is produced based on information that the user inputs, which consists of prompts (typically short instructional texts).

Examples of generative AI tools include chatbots such as ChatGPT and Bing Chat; GitHub Copilot, which produces code based on text prompts; and DALL-E, Midjourney and Stable Diffusion, which produce images from text or image prompts. In addition, generative AI models can be fine-tuned, or custom models can be trained and deployed to meet an organization’s needs. [2]

Many generative AI models have been trained on large volumes of data, including publicly available data from the Internet. Based on the training data, these models generate content that is statistically likely in response to a prompt, [3] for example, by predicting the next word in a sentence. Techniques such as human supervision and reinforcement learning can also be applied to further improve the outputs, [3] and users can provide feedback or change their prompt to refine the response. Generative AI can therefore produce content that looks as though a human produced it.

Generative AI can be used to perform or support various tasks including:

- writing and editing documents and emails
- coding tasks, such as debugging and generating templates and common solutions
- summarizing information
- brainstorming
- research, translation and learning
- providing support to clients (for example, answering questions, troubleshooting)

Challenges and concerns

Before federal institutions start using generative AI tools, they must assess and mitigate certain ethical, legal and other risks. For example, these tools can generate inaccurate content; amplify biases; and violate intellectual property, privacy and other laws. Further, some tools may not meet federal privacy and security requirements. When institutions use these tools, they must protect personal information and sensitive data. As well, because these tools generate content that can look as though a human produced it, people might not be able to tell whether they are interacting with a person or a tool. The use of these tools can also affect the skill and judgment of public servants and can have environmental costs.

Generative AI tools rely on models that pose various challenges, including limited transparency and explainability. They also rely on training data that is difficult to access and assess. These challenges stem in part from large model sizes, high volumes of training data, and the proprietary nature of many of the tools. In addition, the outputs of the models are constrained by the prompts and the training data, which may lack context that is not publicly available on the Internet. Training data could also be outdated; for example, ChatGPT is trained on data up to 2021, so it has a limited ability to provide information on events after that. [4] [5] As well, these tools have limitations that reduce their utility for certain purposes; for example, they tend to perform poorly on tasks related to emotion. [6] [7]

Generative AI could also pose risks to the integrity and security of federal institutions, given its potential misuse by threat actors. Federal institutions should be aware of these risks and ensure that the necessary mitigation measures are in place in accordance with the Canadian Centre for Cyber Security’s guidance on generative AI.

Recommended approach

Federal institutions should explore how they could use generative AI tools to support their operations and improve outcomes for Canadians. However, given the challenges and concerns relating to these tools, institutions should assess and mitigate risks and should restrict their use to activities where they can manage the risks effectively. Given the growing adoption of these technologies in different sectors and by the public, their use in government will help keep pace with the evolving digital landscape.

Federal institutions should evaluate the tools for their potential to help employees, not replace them. When deciding whether to use generative AI tools, public servants should refer to the guide to ethical decision-making in section 6 of *Values Alive: A Discussion Guide to the "Values and Ethics Code for the Public Sector."*

To maintain public trust and ensure the responsible use of generative AI tools, federal institutions should align with the "FASTER" principles:

- **Fair:** ensure that content from these tools does not include or amplify biases and that it complies with human rights, accessibility, and procedural and substantive fairness obligations
- **Accountable:** take responsibility for the content generated by these tools. This includes making sure it is factual, legal, ethical, and compliant with the terms of use
- **Secure:** ensure that the infrastructure and tools are appropriate for the security classification of the information and that privacy and personal information are protected
- **Transparent:** identify content that has been produced using generative AI; notify users that they are interacting with an AI tool;

document decisions and be able to provide explanations if tools are used to support decision-making
- **Educated:** learn about the strengths, limitations and responsible use of the tools; learn how to create effective prompts and to identify potential weaknesses in the outputs
- **Relevant:** make sure the use of generative AI tools supports user and organizational needs and contributes to improved outcomes for Canadians; identify appropriate tools for the task; AI tools aren't the best choice in every situation

For assistance in determining the appropriate use of these tools, public servants should engage with relevant stakeholders such as their institution's legal services, privacy and security experts, Chief Information Office, Chief Data Office and diversity and inclusion specialists. As well, the Canadian Centre for Cyber Security, Statistics Canada and the Office of the Chief Information Officer at TBS are also available to support federal institutions in the responsible use of these tools.

Policy considerations and best practices

Does the *Directive on Automated Decision-Making* apply?

The *Directive on Automated Decision-Making* applies to automated systems, including those that rely on AI, used to influence or make administrative decisions. Like other AI systems, generative AI systems have capabilities that allow them to make assessments or determinations about clients in service delivery. For example, a generative AI system could be used to summarize a client's data or to determine whether they are eligible for a service. [8] These administrative uses have the potential to affect how an officer views and decides on a case, which has implications for the client's rights, interests or privileges. The directive therefore applies to the use of generative AI systems to make or inform administrative decisions.

However, generative AI may not be suited for use in administrative decision-making at this stage. The design and functioning of generative models can limit federal institutions' ability to ensure transparency, accountability and fairness in decisions made by generative AI systems or informed by their outputs. As well, the terms of use for the generative AI products of many leading technology companies prohibit using their products to make high-impact decisions. For example, OpenAI instructs users not to employ ChatGPT in decisions about credit, employment, educational institutions, or public assistance services; law enforcement and criminal justice; and migration and asylum. [9] Similarly, Google prohibits users of their generative AI product from making "automated decisions in domains that affect material or individual rights or well-being." [10] These limitations underscore the importance of complying with the directive's requirement to consult legal services during the design phase of an automation project. The consultation allows federal institutions to understand the legal risks of administrative uses of generative AI systems both for themselves and for their clients.

Not all uses of generative AI are subject to the directive. For example, using generative tools in research or to brainstorm, plan, or draft routine correspondence falls outside the scope of the directive. However, such non-administrative uses are still subject to the laws and policies that govern federal institutions.

Privacy considerations

As with any online system, personal information should not be entered into a generative AI tool or service unless a contract is in place with the supplier and covers how the information will be used and protected. Before using a generative AI tool, federal institutions must also make sure that the collection and use of personal information, including information used to train the tool, meets their privacy obligations.

All personal information used by, created or obtained through, and disclosed for the use of generative AI by federal institutions is subject to the requirements of the *Privacy Act* and related policy instruments. This means that:

- personal information can only be collected if it is directly related to the program or activity
- it may only be used for the purpose for which it was collected or for a use consistent with that purpose
- it has limited permissible disclosures outlined in the legislation
- institutions must be transparent about how they treat and safeguard the personal information they collect once it is under the control of the government

The privacy risks will vary based on how the AI tool collects and processes information about individuals and, potentially, makes decisions about them. An AI tool could, for example, decide whether someone is eligible for a service, determine the level of benefit someone is entitled to, or process survey data to inform policy direction.

The *Privacy Act* requires that a government institution take all reasonable steps to ensure that personal information that is used for an administrative purpose by the institution is as accurate, up-to-date and complete as possible. When using a generative AI system to make or inform decisions about individuals, federal institutions must have confidence that the personal information the system collects, creates or uses is accurate. For this reason, direct collection from the individual is required in most situations. Direct collection also allows for the individual to be notified of the collection and of how their information will be used and managed.

If the output of a generative AI tool results in the creation of new personal information, the new information must also be managed according to privacy requirements. For example, a summary of an application for a service or benefit produced by a generative AI tool could constitute new personal information. Users should validate any personal information created by a generative AI tool to make sure that it is accurate, up-to-date and complete. As well, users must ensure that any new personal information is not disclosed for a purpose that is inconsistent with that for which it was collected. From the example above, sharing the new information about the individual with a different program for an unrelated benefit may not be appropriate and may constitute a privacy breach.

Federal institutions must also make sure that all personal information they collect and use can be made available to the individual concerned and that the individual can access and correct it upon request. Federal institutions must retain personal information that is used to make a decision about an individual for at least two years. This gives the individual enough time to exercise their right to access and correct the information. Federal institutions should not hold onto personal information for longer than required. The longer federal institutions hold personal information, the greater the likelihood of a potential privacy breach.

De-identification and the use of synthetic data can help institutions reduce the impact and likelihood of privacy breaches when training, using and evaluating the outputs of generative AI tools. Privacy Implementation Notice 2023-01: De-identification contains more information about these privacy preserving techniques. Other safeguards such as administrative controls, access rights, and auditing are also important to reduce the risk of inadvertent disclosure or unauthorized access, re-identification or inference, and to generally preserve the privacy of individuals.

Before considering procuring, using or deploying generative AI tools, federal institutions' privacy officials must determine whether a Privacy Impact Assessment is needed.

When federal institutions are building IT solutions that use generative AI, they must make sure they meet privacy requirements. *The Digital Privacy Playbook* contains more information on these requirements and on how to incorporate privacy guidance into IT solutions that use generative AI.

Potential issues and best practices

The following section provides a brief overview of several areas of risk and sets out best practices for the responsible use of generative AI in federal institutions. In addition to the best practices identified for all users of generative AI in the federal government, best practices specific to federal institutions developing or deploying these tools are also identified to ensure that risks are appropriately assessed and mitigated, and to distinguish between the responsibilities of users and developers.

Protection of information

Issue: some generative AI tools do not meet government information security

requirements

The protection of personal, classified, protected and proprietary information is critical when using generative AI systems. The providers of some generative AI tools may inspect input data or use this data to further train their models, which could result in privacy and security breaches. Risks can also arise from input data being stored on servers not controlled by the GC, where data might be retained for longer than necessary, made accessible, further distributed, or vulnerable to a data breach. [11] Some tools, public or otherwise, may not meet privacy and security requirements established in federal law and policy.

Best practices for all users of generative AI in federal institutions

- Don't enter sensitive or personal information into any tools not managed by the GC.
- Don't submit queries on non-GC managed tools that could undermine public trust if they were disclosed. Refer to Appendix B of the *Directive on Service and Digital* for examples of unacceptable uses.
- Understand how a system uses input data (for example, whether it's used as training data or accessible to providers).
- Ask legal services and the departmental chief security officer (CSO) to review a supplier's terms of use, privacy policy and other legal documents before using any system to process sensitive or proprietary information.
- Use infrastructure and tools that are appropriate for the security classification of the information, in accordance with the *Directive on Security Management*.
- Consult the departmental CSO before using, procuring or deploying generative AI for protected or other sensitive information.
- Consider the requirements for information and data residency in the

Directive on Service and Digital and the related guidance in the *Guideline on Service and Digital*.
- Use the “opt-out” feature, where possible, to ensure that prompts are not used to train or further develop an AI system.

Additional best practices for federal institutions deploying a generative AI tool

- Conduct regular system testing prior to and throughout the operation of a system to ensure that it meets key performance targets.
- Plan independent audits for assessing generative AI systems against risk and impact frameworks.

Bias

Issue: generated content may amplify biases or other harmful ideas that are dominant in the training data

Generative AI tools can produce content that is discriminatory or not representative, or that includes biases or stereotypes (for example, biases relating to multiple and intersecting identity factors such as gender, race and ethnicity). [12] [13] [14] Many generative models are trained on large amounts of data from the Internet, which is often the source of these biases. For example, training data is likely to reflect predominant historical biases and may not include perspectives that are less prevalent in the data or that have emerged since the model was trained. [12] Other sources that may contribute to biased content include data filtering, which can amplify the biases in the original training set, [15] framing of the prompt, [16] and model bias. Widespread use of these technologies could amplify or reinforce these biases and dominant viewpoints, and lead to less diversity in ideas, perspectives and language, [12] [17] as well as potential harms.

Best practices for all users of generative AI in federal institutions

- Review generated content to ensure that it aligns with GC commitments, values and ethics and meets legal obligations. This includes assessing for biases or stereotypical associations.
- Formulate prompts to generate content that provides holistic perspectives and minimizes biases.
- Strive to understand the data that was used to train the tool, for example, where it came from, what it includes, and how it was selected and prepared.
- Learn about bias, diversity, inclusion, anti-racism, and values and ethics to improve your ability to identify biased or discriminatory content.
- Notify recipients when content has been produced by generative AI.

Additional best practices for federal institutions deploying a generative AI tool

- Consider potential biases and mitigation approaches from the planning and design stage, including by completing a gender-based analysis plus (GBA Plus) to understand how your deployment of generative AI tools might impact different population groups.
- Consult GBA Plus and other diversity and inclusion experts in your organization to identify impacts of the use of generative AI tools on different population groups and to develop measures to address those impacts.
- Test for biases in the data, model and outputs before deploying a system, and on an ongoing basis.

Quality

Issue: generated content may be inaccurate, incoherent or incomplete

Generative AI technologies can produce content that appears to be well developed, credible and reasonable but that is in fact inaccurate, nonsensical or inconsistent with source data. [18] [19] This content is sometimes referred to as a "hallucination." Also, content generated by AI tools may not provide a holistic view of an issue. Instead, it may focus on prevalent perspectives in the training data. [12] It also might be out of date, depending on the time period the training data covers and whether the system has live access to recent data. The quality of the tools and outputs in different languages should also be considered to ensure compliance with official languages requirements.

The risks associated with inaccurate content will vary based on the context and should be assessed. For example, using generative AI tools to learn about a topic may produce incorrect information or non-existent sources, [20] which, if used in decision-making, could lead to unfair treatment of individuals or misguided policy. As well, the use of generative AI tools for public-facing communications could result in the government sharing inaccurate information, which would contribute to misinformation and erode public trust.

Best practices for all users of generative AI in federal institutions

- Clearly indicate that you have used generative AI to develop content.
- Don't consider generated content as authoritative. Review it for factual and contextual accuracy by, for example, checking it against information from trusted sources.
- Review personal information created using generative AI to ensure it is accurate, up-to-date and complete.
- Assess the impact of inaccurate outputs. Don't use generative AI when factual accuracy or data integrity is needed.

- Strive to understand the quality and source of training data.
- Consider your ability to identify inaccurate content before you use generative AI. Don’t use it if you can’t confirm the content quality.
- Learn how to create effective prompts and provide feedback to refine outputs to minimize the generation of inaccurate content.

Additional best practices for federal institutions deploying a generative AI tool

- Make sure the quality of tools and outputs meets official languages requirements before deployment.
- Notify users that they are interacting with generative AI.
- Use watermarks to help users identify content generated by AI.
- When content is generated by AI, provide links to authoritative sources and encourage users to verify the content at the links provided.
- Provide information about the source of training data and how models were developed.

Public servant autonomy

Issue: overreliance on AI can unduly interfere with judgment, stifle creativity and erode workforce capabilities

Overreliance on generative AI tools can interfere with individual autonomy and judgment. For example, some users may be prone to uncritically accept system recommendations or other outputs, which could be incorrect. [21] [22] Overreliance on the system can be a sign of automation bias, which is a tendency to favour results generated by automated systems, even in the presence of contrary information from non-automated sources. [21] As well, confirmation bias can contribute to overreliance [21] because the outputs of generative AI systems can reinforce users' preconceptions, especially when prompts are written in a way that reflects the user's assumptions and beliefs. [23] Overreliance on AI systems can result in a decline in critical thinking and can limit diversity in thought, thereby stifling creativity and innovation and resulting in partial or incomplete analyses. Overreliance on AI can impede employees' ability to build and maintain the skills they need to complete tasks that are assigned to generative AI systems. This could reinforce the government's reliance on AI and potentially erode workforce capabilities.

Best practices for all users of generative AI in federal institutions

- Consider whether you **need** to use generative AI to meet user and organizational needs.
- Consider the abilities and limits of generative AI when assigning tasks and reviewing system outputs.
- Build your AI literacy so that you can critically assess these tools and their outputs.
- Use generative AI tools as aids, not as substitutes. Do not outsource a skill that you do not understand or possess.
- Form your own views before you seek ideas or recommendations from AI tools.
- Learn how to write prompts that are likely to result in content that provides a holistic perspective and minimizes biases.
- Always review content generated by AI, even if the system seems to be reliable in providing accurate responses.

Legal risks

Issue: generative AI poses risks to human rights, privacy, intellectual property protection, and procedural fairness

The government’s use of generative AI systems poses risks to the legal rights and obligations of federal institutions and their clients. These risks arise from the data used to train AI models, the way systems process input data, and the quality of system outputs.

The use by suppliers or federal institutions of copyright-protected materials like articles, books, code, paintings or music to train AI models may infringe intellectual property rights. The use or reproduction of the outputs generated by these models could also infringe on such rights if they contain material that is identical or substantially similar to a copyright-protected work. Further, the ownership of content created by or with the help of generative AI is uncertain. Privacy rights could also be at risk because data used to train generative AI models could include unlawfully collected or used personal information, including personal information obtained from publicly accessible online sources.

Risks could also arise from the opacity of generative AI models and their potential for producing inaccurate, biased or inconsistent outputs. This opacity makes it difficult to trace and understand how the AI system produces outputs, which can undermine procedural fairness in instances where a federal institution is obliged to provide clients with reasons for administrative decisions, such as decisions to deny benefits. The quality of AI outputs can also impact individuals’ legal rights. For example, biased outputs could lead to discrimination in services, potentially violating human rights.

These risks extend beyond decision-making scenarios. When federal institutions use generative AI tools to help the public find information (as is the case, for example, with the use of chatbots on departmental websites) or to produce public communications, there’s a risk that these tools will generate inappropriate content or misinformation that could contribute to or cause harm for which the government could be liable.

Best practices for all users of generative AI in federal institutions

- Consult your institution’s legal services about the legal risks of deploying generative AI tools or using them in service delivery. The consultation could involve a review of the supplier’s terms of use, copyright policy, privacy policy and other legal documents.
- Comply with the *Directive on Automated Decision-Making* when using generative AI in administrative decision-making.
- Check whether system outputs are identical or substantially similar to copyright-protected material. Give proper attribution, where appropriate, or remove this material to minimize the risk of infringement of intellectual property rights.
- Consult designated officials on the licensing and administration of Crown copyright if you are planning to include outputs in public communications, in accordance with the *Procedures for Publishing*.
- Evaluate the quality of outputs for factual inaccuracies, biases or harmful ideas that may conflict with GC values.
- Keep up-to-date on legal and policy developments related to AI regulation.

Additional best practices for federal institutions deploying a generative AI tool

- Verify the legality of the method used to obtain data for training AI models and make sure you have permission to use the data for this purpose. Where feasible, train your model using open-source data that has no restrictions on such use.
- Be transparent about your use of generative AI, including by notifying users if they are interacting with a system rather than a human. Where relevant, include a disclaimer to minimize liability

risks.
- Use watermarks to help users identify generated content.

Distinguishing humans from machines

Issue: people may not know that they are interacting with an AI system, or they may wrongly assume that AI is being used

Conversational agents or chatbots that use generative AI can produce responses that are so human-like that it may be difficult to distinguish them from those of a real person. [24] As a result, clients may be misled into believing that they are interacting with a human. Similarly, clients might assume that an email they have received was written by a person when it was actually generated by an AI tool. On the other hand, clients might think they are interacting with an AI tool when they are actually dealing with a real person. Transparency about whether a client is interacting with a person or a chatbot is essential to ensure that the client is not misled and to maintain trust in government.

Best practices for all users of generative AI in federal institutions

- Clearly communicate when and how the GC is using AI in interactions with the public.
- Inform users when messages addressed to them are generated by AI.

Additional best practices for federal institutions deploying a generative AI tool

- Consider offering non-automated means of communicating with the GC.
- Use watermarks so that users can identify content generated by AI.
- Publish information about the system, such as a plain-language description of how it works, the reasons for using it, and the quality

assurance steps taken.

Environmental impacts

Issue: the development and use of generative AI systems can have significant environmental costs

The development and use of generative AI systems can be a significant source of greenhouse gas emissions. These emissions come not only from the compute used to train and operate generative models but also from the production and transportation of the servers that support the AI programs. [25] While generative AI has the potential to help combat climate change, its use must be balanced against the need for swift and drastic action to reduce global greenhouse gas emissions and avert irreversible damage to the environment. [26]

Best practices for all users of generative AI in federal institutions

- Use generative AI tools hosted in zero-emission data centres.
- Use generative AI tools only when relevant to program objectives and desired outcomes.

Additional best practices for federal institutions deploying a generative AI tool

- Consider whether your AI supplier has set any greenhouse-gas reduction targets. [27]
- Complete an environmental impact assessment as part of the proposal to develop or procure generative AI tools. Make sure any decision to procure these tools is made in accordance with the *Policy on Green Procurement*.

Use of this guide and additional support

available

As departments further evolve their guidance on use of generative AI, this document is to be used as overarching guidance to build from. For more information, including guidance on specific uses of generative AI, contact the TBS Responsible Data and AI team (ai-ia@tbs-sct.gc.ca). Additional resources exist within the federal government which institutions can access by contacting the Communications Security Establishment (including Canadian Centre for Cyber Security's guidance on generative AI) and Statistics Canada. The community of practice and the TBS guide will continue to evolve over the next number of years.

Frequently asked questions

▼ Can I use generative AI to draft emails or briefing notes?

Yes. Depending on the context, you can use a generative AI tool to support drafting of emails or briefing notes that don't contain personal or sensitive information. The person generating the content is responsible for making sure that:

- input data does not include protected, classified or other sensitive information
- generated content is accurate, non-partisan, unbiased, and doesn't violate intellectual property laws
- management is notified that a generative tool was used in the development of the product

▼ Can I use generative AI to develop content for public communications (for example, web posts, social media)?

Use caution. When you generate content, you are responsible for making sure it is accurate, clear, non-partisan and unbiased. You are also responsible for making sure permissions to reproduce, adapt, translate or publish third-party material have been secured and that the content does not violate intellectual property laws. You should also inform the public of any significant use of generative AI in the production of content. It is also critical to ensure that outputs are trusted given the potential reach and impact of public communications.

▼ Can I use generative AI for programming tasks?

Yes, but you must consider the security classification of the code. Also, when it comes to code generation, some generative AI tools can produce content that violates the open-source licences of the source code they were trained on. To address this issue, use available tools to identify potential matches in public code repositories or limit the use of generative AI to tasks like debugging or code explanation.

▼ Can I use generative AI to inform policy?

Yes, but you must be mindful of the strengths and limits of generative AI tools and tailor the tasks you assign to them accordingly. You can use these tools to assist with research during policy development, but don’t use them to recommend, make or interpret policy.

When deciding on policy positions, make your own value judgments, in consultation with the relevant stakeholders and consistent with applicable laws. Strive to be transparent and vigilant about any significant use of generative AI during the policy process, including in research and stakeholder engagement. The prompts used in such contexts should not include any information that would pose legal or reputational risks to the government.

▼ Can I use generative AI to automate assessments, recommendations or decisions about clients?

Use caution when considering whether to use generative AI in administrative decision-making. Carefully consider how you will comply with the *Directive on Automated Decision-Making*, which seeks to ensure transparency, accountability and fairness in decisions made or informed by automated systems such as those that use generative AI. For example, make sure that you understand how the tool produces its outputs and that you can find the data it relied on. You should assess outputs for factual accuracy and undue bias toward clients. You should also consider potential variation in outputs produced in response to similar prompts, which could lead to inequalities in the treatment of clients.

Bibliography

1 McKinsey & Company, "What is Generative AI?," 19 January 2023. [Accessed 8 May 2023].

2 F. Candelon, A. Gupta, L. Krayer and L. Zhukov, "The CEO’s Guide to the Generative AI Revolution," 7 March 2023. [Accessed 11 May 2023].

3 K. Martineau, "What Is Generative AI?," 20 April 2023 [Accessed 8 May 2023].

4 OpenAI, "Documentation - Models Overview," [Accessed 8 May 2023].

5 OpenAI, "What is ChatGPT?," [Accessed 8 May 2023].

6 J. Kocoń, I. Cichecki, O. Kaszyca, M. Kochanek, D. Szydło, J. Baran, J. Bielaniewicz, M. Gruza, A. Janz, K. Kanclerz, A. Kocoń, B. Koptyra, W. Mieleszczenko-Kowszewicz, P. Miłkowski, M. Oleksy, M. Piasecki, Ł. Radliński, K. Wojtasik, S. Woźniak and P. Kazienko, "ChatGPT: Jack of All Trades, Master of None," *SSRN preprint,* pp. 1-40, 28 February 2023.

7 K. Yang, S. Ji, T. Zhang, Q. Xie and S. Ananiadou, "On the Evaluations of ChatGPT and Emotion-enhanced Prompting for Mental Health Analysis," *arXiv preprint,* 2023.

8 Council of the European Union, "ChatGPT in the Public Sector - Overhyped or Overlooked?," European Union, 2023.

9 OpenAI, "OpenAI Usage Policies," OpenAI, 2023 [Accessed 12 May 2023].

10 Google, "Google Generative AI Prohibited Use Policy," Google, 2023 [Accessed 12 May 2023].

[11] OpenAI, "March 20 ChatGPT Outage: Here’s What Happened," 24 March 2023 [Accessed 10 May 2023].

[12] E. M. Bender, T. Gebru, A. McMillan-Major and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜," *FAccT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency,* pp. 610-623, 2021.

[13] A. S. Luccioni, C. Akiki, M. Mitchell and Y. Jernite, "Stable Bias: Analyzing Societal Representations in Diffusion Models," *arXiv preprint,* pp. 1-44, 2023.

[14] W. Guo and A. Caliskan, "Detecting Emergent Intersectional Biases: Contextualized Word Embeddings Contain a Distribution of Human-like Biases," *AIES '21: Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society,* pp. 122-133, July 2021.

[15] OpenAI, "DALL·E 2 Pre-training Mitigations," 28 June 2022 [Accessed 4 May 2023].

[16] University of Waterloo Library, "ChatGPT and Generative Artificial Intelligence (AI): Potential for Bias Based on Prompt," 27 April 2023 [Accessed 5 May 2023].

[17] C. Bjork, "ChatGPT Threatens Language Diversity. More Needs to Be Done to Protect Our Differences in the Age of AI," 9 February 2023 [Accessed 5 May 2023].

18 Z. Ji, N. Lee, R. Frieske, T. Yu, D. Su, Y. Xu, E. Ishii, Y. Bang, W. Dai, A. Madotto and P. Fung, "Survey of Hallucination in Natural Language Generation," *arXiv preprint,* pp. 1-47, 7 November 2022.

19 OpenAI, "Introducing ChatGPT," 22 November 2022 [Accessed 1 May 2023].

20 H. Alkaissi and S. I. McFarlane, "Artificial Hallucinations in ChatGPT: Implications in Scientific Writing," *Cureus,* vol. 15, no. 2, p. e35179, 2023.

21 S. Passi and M. Vorvoreanu, "Overreliance on AI: Literature review," June 2022. [Accessed 3 May 2023].

22 Z. Buçinca, M. B. Malaya and K. Z. Gajos, "To Trust or to Think: Cognitive Forcing Functions Can Reduce Overreliance on AI in AI-assisted Decision-making," *Proceedings of the ACM on Human-Computer Interaction,* vol. 5, no. CSCW1, pp. 1-21, April 2021.

23 M. Grawitch, "Confirmation Bias in the Era of Large AI," 1 May 2023 [Accessed 11 May 2023].

24 A. James, "ChatGPT Has Passed the Turing Test and if You're Freaked Out, You're Not Alone," 29 March 2023 [Accessed 4 May 2023].

25 S. McLean, "The Environmental Impact of ChatGPT: A Call for Sustainable Practices In AI Development," 28 April 2023 [Accessed 4 May 2023].

[26] IPCC, "Summary for Policymakers. In: Global Warming of 1.5°C. An IPCC Special Report on the Impacts of Global Warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty," Cambridge University Press, Cambridge, UK; New York, NY, USA, 2018.

[27] United Nations Climate Change, "Race To Zero Campaign," [Accessed 5 May 2023].

Date modified:

2023-09-06