Guidelines for Government of Saskatchewan Employees in Using Generative Artificial Intelligence

Purpose:

The purpose of this document is to provide guidance and good practices for Government of Saskatchewan (GOS) employees in the use of generative artificial intelligence (AI) tools in their day-to-day work. It is important for employees to be aware of generative AI-related issues and risks so they can manage information securely for all GOS information systems and limit data use and sharing on third-party applications not authorized by GOS. This document serves as a guide for GOS employees on how to use generative AI tools safely and securely.

Scope:

These guidelines deal with several types of generative AI, including but not limited to written (e.g., ChatGPT), visual (e.g., Dall-E 2, Bing Image Creator, Midjourney) or audio (e.g., Bommy, Soundraw) and are applicable to all GOS employees' work.

The definition of terms used in this document can be found in Appendix A.

The Government of Saskatchewan's existing policies apply to online activities as fully as they do to activities in any other circumstance or venue. Before using any type of generative AI, employees must review the policies listed in Appendix B.

Prior to using Generative AI

Three things to keep in mind prior to using generative AI:

- **Generative AI is not just a product you are using:** When you use generative AI, you become a contributor as well as a user. Unlike traditional search engines like Google, generative AI captures and keeps all information entered into it. The tool uses inputs to add to its database and then uses that data to produce results for other users, which means the information entered is now publicly available.
- **Generative AI does not assess truthfulness or accuracy in its output:** As mentioned above, generative AI uses data from all its users. This data can be true or false, comprised of pieces of information from various sources and the tool does not differentiate or provide a necessarily complete answer. GOS employees need to use critical thinking to validate the results, as they are responsible and accountable for the outcomes.

Generative AI contains inherent bias: Many assume that generative AI tools are free of bias. However, the humans who created the tools have biases (conscious or unconscious), which can be written into or learned by the tool. If biases exist in the tool, they will be present in the final product. It is important to review the final product for bias, as biases do not comply with ethical policies and regulations.

Should I use Generative AI?

Prior to using generative AI, it is important to identify potential risks, such as who and what may be at risk. The table below provides some examples to consider.

Who and What are at risk? *	
Who	**What**
Citizens of Saskatchewan	Privacy and information security
Government of Saskatchewan	Public trust/transparency
Ministries, Agencies, and Crowns	Biases and ethical considerations
Government of Saskatchewan Employees	Control and validity
Consultants and 3rd party contracts	
	***Further information is located in Appendix B**

Once you've identified who and what could be at risk by using generative AI, the questions below will assist you in determining if you should proceed with using generative AI. The questions are intended to help the user consider the impact of who and what is at risk when using generative AI.

What are you trying to accomplish?

Generative AI is a useful tool for certain kinds of work, but should not be used for all types of work. Generative AI is best used as a tool to help generate ideas for a new project, conduct an inter-jurisdictional scan, create a first draft, or brainstorm. It should not be used to create final products such as briefing notes or decision-making documents (e.g., Cabinet documents). Generative AI is a tool that can help you at your job, but it is not meant to do the work for you. The following checklists are not exhaustive but are examples of various types of work:

Continue if you want to:

☐ Get ideas.
☐ Produce a first draft to work from.
☐ Brainstorm.
☐ Conduct initial research on an unfamiliar policy/topic (e.g., inter-jurisdictional scan).
☐ Summarize long publicly available articles or reports.
☐ Figure out wording for something.
☐ Get started when stuck.

Pause to assess further if you want to:

☐ Write a decision-making document.
☐ Write a letter to a citizen.
☐ Write a business case or plan.
☐ Have the generative AI tool do most of the work for you.

Is what you're going to put into the generative AI tool private or confidential?

Information you put into a generative AI tool becomes public and part of the tool's resources for others to use. Information that is already publicly available is okay to use as input. Any data or information that is not publicly available should be considered confidential and not used in a generative AI prompt or query. A good rule of thumb is that if the information is available on a public website, then it can be considered public information and can be used.

If information is publicly available, **please proceed.**

If information is not publicly available, **please pause to assess.**

If information can, in any way, be considered private or confidential, **DO NOT PROCEED** using generative AI.

Did you ask for and check the sources used by the generative AI tool?

Generative AI is not like Google in that it does not search for sources that you get to look further into. It uses information that other users have input to produce results for your query. It does not assess the accuracy of information in its output. Asking for sources or checking information against other sources is your responsibility to ensure the accuracy of the output. Citing your use of generative AI in your final product is best practice, so others are aware of your source.

If the tool cannot provide sources or you cannot verify sources, **pause to assess** whether to continue.

If you cannot verify results with some other source, **BE VERY CAUTIOUS** with the results.

Are you willing to be transparent about your source?

Citing a generative AI tool as a source is important in being transparent about a document's origins. Be clear about what role the output played in the final product. For example, if you generated a first draft of a document, cite the generative AI tool as a source for the first draft, which was then rigorously reviewed and edited. If you can explain how the tool came to the decision it made and how those decisions impacted citizens, this will maintain trust with citizens. The overall generative AI process needs to be transparent, openly communicated, and explainable to those directly and indirectly affected.

If you're still unsure about using generative AI in your particular situation, please refer to the Risk Assessment in Appendix C for further analysis and detailed examples.

Best Practices When Using Generative AI

Below are some best practices that should be adhered to when using generative AI tools:

- **Never** use sensitive information or personal data in these tools. Beyond existing data protection laws, the government has no oversight over how data entered into web-based generative AI tools is then used. Therefore, you should not put information into generative AI tools that, if compromised or lost, could have damaging consequences for individuals, groups, an organization, or the government more generally. Personal information should not be entered into a generative AI tool. A general rule is if the information is not publicly available, it should not be entered into a generative AI tool. For example, asking ChatGPT to draft a press release on a new government policy is not advisable because that information is not available to the public.
- **All employees should be aware of the benefits and challenges of using generative AI:** Employees using generative AI need to be aware of its challenges and shortcomings to ensure they produce trustworthy, fair and unbiased results. Prior to using generative AI, discuss the possible use case with your supervisor.
- **Clearly document your process.** The public needs to be aware that there is a process in place for how employees use generative AI and specific examples of what it is to be used for, so they can be confident in the work produced. Part of this process is to go through the questions at the beginning of this document.
- **Reviewing the validity of the output:** It is important to validate, or fact-check, the information produced by the generative AI tool. This can be accomplished by asking the tool for its sources and reviewing them.

Reporting & Review:

It is recommended that an employee disclose when, where, and how generative AI was used for a specific project. This can be achieved by adding a disclaimer to the final product, letting your manager know generative AI was used, and providing the answers to the four questions at the beginning of the guidelines when required.

It is recommended that the guidelines be reviewed every six months to ensure they best reflect the current advances in the field, including emerging best practices and offering a better understanding of the use cases for generative AI. If there is a major development in the generative AI field, then the guidelines should be reviewed accordingly.

Contact Information:

Jennifer Block, A/Executive Director
Citizen Centric Delivery Program Branch
jennifer.block@gov.sk.ca

Appendix A: Definitions

Bias: An inclination of prejudice towards or against a person, object, or position. It does not necessarily relate to human bias; it can arise, for example, through the limited context in which a system is used.

Explainability: The ability to explain both the technical processes of the generative AI system and the reasoning behind the decisions or predictions the system makes, explaining why a generative AI system reached a particular decision, recommendation, or prediction. Explainability is crucial for building and maintaining users' trust and verifying the information as it allows human users to comprehend and trust the results and outputs.

Generative Artificial Intelligence: A set of relatively new technologies that leverage very large volumes of data along with some machine learning techniques to produce content based on inputs (known as prompts) from the users. A user can use this tool by asking a question on any subject to get a detailed answer. The user can then interact with the system about the answer given. Generative AI tools can also summarize articles and can write computer code. For example, ChatGPT is a web-based, generative AI tool that is readily available on the internet (a paid version does exist).

Machine Learning: The use and development of computer systems that are able to learn and adapt without following explicit instructions by using algorithms and statistical models to analyze and draw inferences from patterns in data.

Appendix B: Existing Government of Saskatchewan Resources

The Government of Saskatchewan's existing policies apply to online activities as fully as they do to activities in any other circumstance or venue. Prior to using any type of generative AI, employees and managers should review the policies listed below:

1. Access and Privacy
2. Security
3. Asset Management Policy
4. Security Compliance Policy
5. Information Protection Security Control (IPSC) for Classified Data
6. User Acceptable Use Policy

Appendix C: Risk Assessment

Possible Risks

There are many different types of risk to consider prior to using generative AI. Below is a list of possible risks employees could consider when determining to use generative AI:

1. **Privacy and Information Security:** Generative AI could breach privacy legislation (*The Freedom of Information and Protection of Privacy Act (FOIP)* and *The Health Information Protection Act (HIPA)* and expose government data, including citizens' data, to malicious actors. Every employee of GOS has a legislated duty to protect personal information collected, used, and disclosed. Privacy breaches lead to harm, such as reputation, humiliation, identity theft, physical harm, etc. These harms directly affect the public's trust in

government.

a. **Things to consider when evaluating this type of risk include:**
    i. Adhere to the Government of Saskatchewan's privacy legislation (FOIP and HIPA).
    ii. Refer to the Government of Saskatchewan's Information Security Policies and User Acceptable Use Policy.
    iii. Avoid using information that could breach privacy legislation. Potential questions to ask: 1. Can an individual be identified from the information? If yes, is the individual's identity reasonably ascertainable? 2. Is the information obviously about the individual? Does the information reveal a fact or opinion about the individual? Refer to the Information Protection Security Controls for Classified Data (IPSC) for additional information.
    iv. Adhere to the Government of Saskatchewan's Cybersecurity policy. Refer to the IT Security Handbook

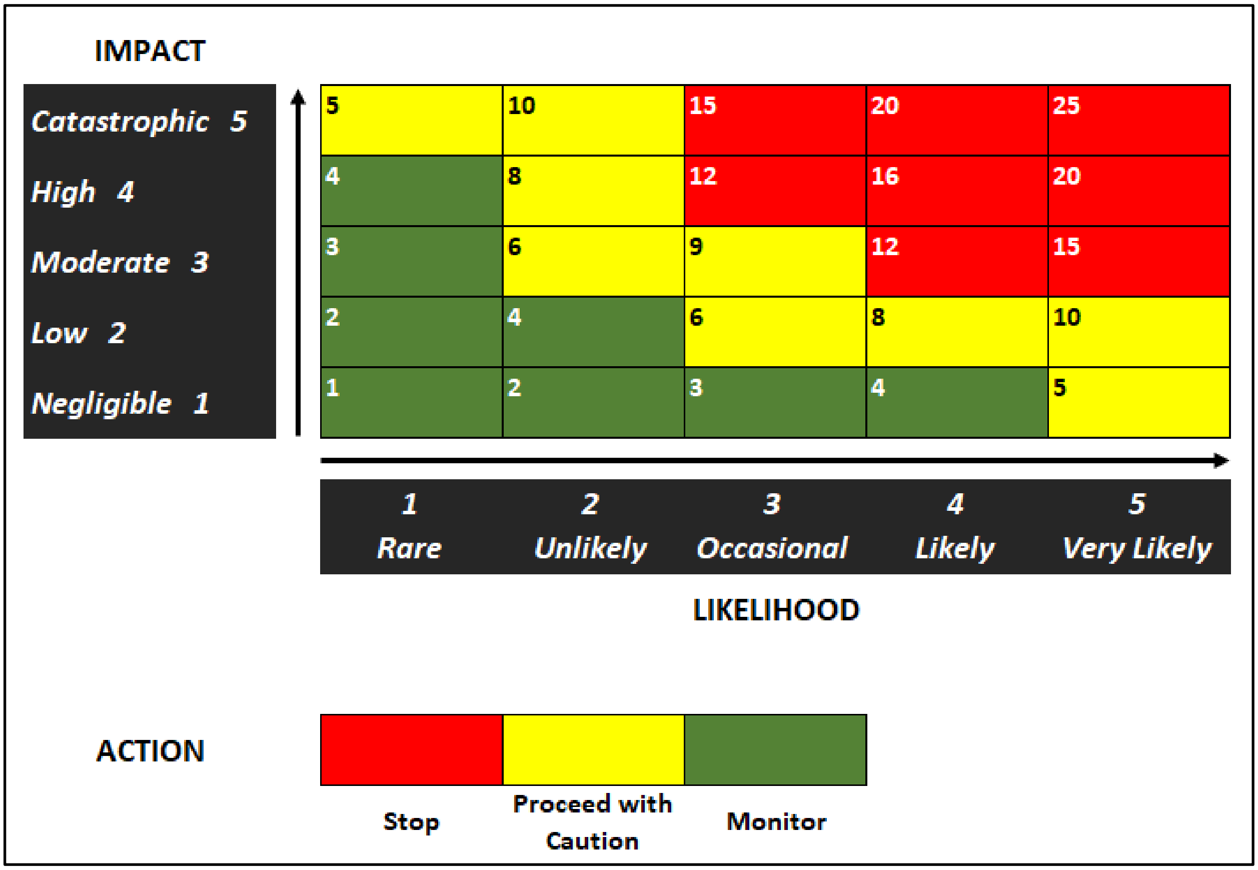
2. **Public Trust:** Currently, there is no verifiable governance regarding data in generative AI systems. When data are entered into generative AI, they become public and could pose a threat to the confidentiality of that data. Any data introduced to a generative AI system can be used by that system to produce results for other users. GOS employees cannot depend on the system to keep confidential information confidential. The unverifiability of data in the system also means that its outputs are not necessarily reliable and must be verified. Both these threats, if discovered by the public, could erode trust in the government's ability to handle confidential information and in its ability to produce quality, accurate work.
    a. Things to consider when evaluating this type of risk include:
        i. Avoid using generative AI outputs directly without assessing to check for:
            1. Fallacy;
            2. Bias; and
            3. Plagiarism
        ii. Use professional due diligence to ensure accuracy.

3. **Bias & Ethical Considerations:** The output from generative AI could be biased and might not align with GOS's mandate.
    a. Things to consider when evaluating this type of risk include:
        i. Determine if the output generated by AI aligns with GOS's mandate and ethical values.
        ii. Verify the sources used by the tool. Some sources have known biases. Some sources are known to be trustworthy.

4. **Control and Validity**: The GOS does not have the ability to control generative AI and its content. The data used by generative AI systems is ever-expanding. Generative AI developers have their own priorities regarding the data they provide to their system, which has a significant impact on output. The version of generative AI being used might be limited due to data model training. For example, ChatGPT version 3.5 (free) only has data up to the year 2021, and only the paid version has current data.
    a. **Things to consider when evaluating this type of risk include:**
        i. Employees are required to understand the limitations of generative AI generally and to assess, as much as possible, the system's approach to data gathering.

Risk Matrix

The risk matrix below is designed to guide an employee's risk assessment of using generative AI. This risk matrix was developed for these Guidelines specifically with the goal of assisting employees in identifying the risk level of using generative AI.



IMPACT	1 Rare	2 Unlikely	3 Occasional	4 Likely	5 Very Likely
Catastrophic 5	🟨 5	🟨 10	🟥 15	🟥 20	🟥 25
High 4	🟩 4	🟨 8	🟥 12	🟥 16	🟥 20
Moderate 3	🟩 3	🟨 6	🟨 9	🟥 12	🟥 15
Low 2	🟩 2	🟩 4	🟨 6	🟨 8	🟨 10
Negligible 1	🟩 1	🟩 2	🟩 3	🟩 4	🟨 5

LIKELIHOOD

ACTION	🟥	🟨	🟩
	Stop	Proceed with Caution	Monitor

Using a scale (1 to 5) helps to determine the level of Impact and Likelihood (or probability) the risk could bring to the GOS and its citizens. Multiply the score for Impact and Likelihood to arrive at a total risk score, then use the colour legend to identify the impact level it falls under.

Impact x Likelihood = Risk

For example, a scenario with the possibility of low Impact and occasional Likelihood would have a score of 6 (2 x 3 = 6) and fall into the **Proceed with Caution** category. A scenario with high Impact and very probable Likelihood would have a score of 20 (4 x 5 = 20) and fall into the **Stop** category.

When assessing the score:

Stop (red) means generative AI should not be used by the employee, and you should not proceed. The potential risk could result in significant threats or loss that stop business continuity (financial, operational etc.).

Proceed with Caution (yellow) refers to taking steps to mitigate potential risk.

Monitoring (green) means tracking and evaluating impact over time to determine the next steps.

Use examples of risk assessment

The table below provides examples of risk assessment to possible applications of ChatGPT on relevant tasks within the GOS, the actions to take for risk management, and its rationale.

Task	Impact	Likelihood	Risk Score	Action	Rationale
Writing a Legislative Decision Item (LDI) using generative AI	5	4	20	Stop	An Impact score of 5 is selected here because an LDI is a very confidential document that influences government operation and a Likelihood of 4 is selected because some employees may see generative AI as an easy, convenient tool for completing an LDI. Using ChatGPT for this task will involve sharing confidential data at a stage that is not for the public. This catastrophic risk can impact the government's reputation and trust. Therefore, do not use generative AI tools like ChatGPT to write confidential documents like this.
Writing Minister's briefing notes using generative AI	4	4	16	Stop	An Impact score of 4 is selected here because a Minister's briefings have a high impact as they inform decision-making. A Likelihood score of 4 is selected because some employees may see generative AI as an easy, convenient tool for writing a Minister's briefing note. Using ChatGPT for this task is a high risk as ChatGPT uses data models from unknown sources that cannot be validated. This means the generated outcome could be false information that will put the

					minister’s briefing notes a high risk of intellectual rights challenges when presenting the outcome to the public. To avoid this risk, consider writing to the ministers. briefing note yourself.
Using generative AI for an inter-jurisdictional scan	3	3	9	Proceed with Caution	An Impact score of 3 is selected here because an inter-jurisdictional scan is created from existing public documents. A Likelihood of 3 is selected because an inter-jurisdictional scan is a task that occurs specifically to complete a project. Using generative AI like ChatGPT for this task is a moderate to low risk as it involves already existing public information. However, you still need to take action to validate the information source of the outputs. Also, be sure to phrase any generative AI prompts in a way that does not use confidential information.
Writing a job description using generative AI	2	4	8	Monitor	An Impact score of 2 is selected here because job descriptions have a low impact; they contain general information. A Likelihood of 4 is selected because some employees may see generative AI as an easy, convenient tool for writing a job description. Using generative AI for this task is a low risk, considering available information is public data. However, you need to monitor the generated outcome and review for bias and any infringements of fundamental rights.