DPC/G13.1

Guideline for the use of Large Language Model Artificial Intelligence Tools and Utilities

Purpose

This guideline covers the limitations and risks associated with the use of Large Language Model (LLM) Artificial Intelligence (AI) within the South Australian (SA) Government, and outlines guidance to assist SA Government agencies (agencies) when considering the safe usage of this technology in their agency. It also provides some practical do's and don'ts in the interim as long-term implications are considered.

Scope

This guideline applies to:

- All SA Government agencies and personnel operating on behalf of the agencies.
- Non-government suppliers and personnel that access SA Government information and resources.

Guidance detail

Like other forms of AI, large language model (LLM) AI learns how to take actions from past data, based on training. However, LLM AI is unlike other forms of AI, which will simply categorise or identify data. The LLM AI creates brand new content and creates human-like responses based upon a training data set.

LLM technologies (also known as generative AI), such as OpenAI's ChatGPT, Bing AI Chat, and Google Bard, have highlighted the need to consider the safe use of such technologies within the SA Government context. The potential benefits of these tools need to be balanced against the limitations of the products and the risks of their use.

Benefits

LLM AI can automate a wide variety of tasks that may take a human days or weeks to perform, freeing up employees to focus on higher value work and to lift productivity.

The capabilities of these technologies can be applied to nearly all aspects of business. For instance, LLM AI can generate reports, handle customer complaints (simultaneously), create content marketing materials such as email campaigns and social media posts, create and debug code, create presentations, generate ideas, and help with brainstorming.

Risk Management

While the potential of AI applications is extensive and growing, like every technology or tool, SA Government employees must remain responsible for their work and continue to meet all their obligations including those relating to privacy, confidentiality, and integrity, ensuring the security of public sector data and responsible handling of personal information.

Prior to authorising the use of these tools, SA Government agencies should conduct their own risk assessments, balancing potential benefits and risk.

Limitations and risks

There are several major limitations inherent with the use of LLM AI, which must be considered and mitigated for acceptable SA Government use.

Refer to the table below for guideline details.

Limitations and Risks	Mitigating Controls
General While LLM AI tools are still evolving, comprehensive guidance is required to ensure appropriate use within agencies.	• Consider developing an agency strategy for the safe adoption of LLM AI technologies. • Establish an AI governance and compliance framework for agency use. • Properly identify and document the risks associated with the use of AI technologies. • Conduct awareness campaigns for the safe and ethical use of AI tools. • Define standard business use cases based on business needs and security risks.

Information confidentiality
Inputs should automatically be considered as available in the public domain.
Employees can easily expose sensitive and proprietary government data in the questions and prompts they provide to LLM AI tools. Any data provided into current LLM AI tools are potentially stored indefinitely, are likely stored outside of Australia, and may not be stored securely.
Data provided may be used to train third-party models in the future or used to generate responses to external users. Additionally, information provided to these technologies can potentially be viewed by service provider employees and subcontractors, future users of the system and hackers.
	• Instruct employees via clearly understood and widely disseminated policies to not use prompts that expose official government or personal information. For example, such policies may include:
	  ○ Disallow any cut-and-paste of enterprise content, such as emails, reports and chat logs into prompts.
	  ○ Disallow any inputs that include customer data.
	  ○ Remind staff of their obligations under the Code of Ethics for the SA Public Sector and the SA Government Information Privacy Principles.
	• Consider a Data Loss Prevention (DLP) solution to monitor or control the egress information flows which contain sensitive information.
	• Establish a self-hosted AI tenancy that is segregated from public networks and users (refer to Azure Open AI below).

Data integrity and accuracy Currently available LLM AI technologies are trained using immense data sets of internet-sourced data, including text scraped from unverified sources such as Twitter, Wikipedia and Reddit. Subpar training data can lead to responses that are insufficient, obsolete or contain sensitive information and biases, leading to biased, prohibited, or incorrect responses.	• Enforce by policy a manual review model to spot incorrect or misinformed results. • Manually review all model output and use it only as a first draft tool. Inaccurate or biased outputs could lead to liability, reputation loss or harm. • Use in a controlled setting where generated text can be properly evaluated and tuned.

One of the most widespread and hard to spot problems in the AI context are 'hallucinations' – these are confident responses by AI that are not justified by its training data, where it makes up the answer and reports it as fact. It has been shown that AI not only hallucinates, but also can amplify hallucinations. Generally, manufacturers of publicly available AI tools ensure that their terms of use make no warranties as to the accuracy or quality of the outputs, and the data it draws upon may not be current. Agencies must remain accountable for the accuracy of all sources, including AI generated outputs, and should research and verify sources as required before they are used in their work. Agencies should also ensure that any attribution obligations are met, with some AI tool manufacturers requiring users attribute the role it has played in drafting or editing content.	• Implement standard processes for configuration and change control among business operations and systems. • If used for developing or checking code, ensure that secure software development principles are adhered to. • Remain responsible and accountable for your work, including to critically assess all AI generated outputs and validate quality and accuracy with other sources. • Thoroughly understand terms of use and seek legal advice as necessary.
Legal and privacy Copyright: Users should be aware of the potential copyright violations with the use of LLM AI tools. Terms of use usually put the responsibility to identify potential copyright violations in the output on the user. In some instances, manufacturers have also been accused of using copyrighted data for training models. Legal Risks: Given the new and evolving nature of the AI technology, agencies should be mindful of broader legal risks, including in relation to any terms of use that apply. For example, it is usual for the terms of use to require users to indemnify the manufacturer for all losses resulting from misuse. Privacy: The privacy policies and terms of use of these technologies are unlikely to meet the requirements of the SA Government to sufficiently protect personal or government information.	• Seek legal advice and conduct a compliance analysis and document the applicable regulatory and contractual requirements. • Review the privacy policy / terms of use issued by the AI service provider regularly. • Conduct privacy impact analyses on the AI business use cases. • Establish or update the agency's privacy policies to accommodate the use of AI tools to meet applicable obligations - including in relation to integrity, privacy, security, human rights, anti-discrimination, administrative and other laws • Remind staff of their obligations under the Code of Ethics for the SA Public Sector and the SA Government Information Privacy Principles.
Ethics and fairness Relying on vast amounts of algorithms and data carries the inherent risk of bias or discriminatory content in outputs, and the algorithms used to generate outputs are largely unknown to general users.	• Regularly monitor and review use of AI tools to ensure they are being used responsibly ethically, safely, and according to law and government policy. • Establish an incident reporting channel and investigation procedure to identify and handle the misuse of AI tools.

The risk is particularly high where outputs may inform government decision making that have the potential to harm or impact adversely on individuals, and in the case of some manufacturers, the terms of use note that use in government context carries a greater risk of potential harm. Data used for AI technology development and training could be insufficient, obsolete or contain sensitive information and biases, leading to biased, prohibited, or incorrect responses. Given these AI tools can also adopt different personas through text, there is also a potential risk of fraud or misuse: Deepfakes: These outputs generated by AI could appear realistic, but actually be fake content. SA Government employees must be vigilant to identify fake news, misinformation, impersonations, or efforts to manipulate public opinion. Fraud and abuse: Malicious actors may use AI chatbots for writing fake reviews, spamming, phishing, and social engineering. Moreover, researchers have found many ways in which some platforms can aid in the development of malware.	• Be vigilant and report on the identified events that violate the ethics principles. • Consider the potential exploitation and attacks of the use of AI tools from external and internal malicious actors.
Microsoft Azure Open AI To reduce the information confidentiality risk associated with the use of open access LLM AI, agencies should consider using Microsoft's Azure Open AI service. The Microsoft Azure OpenAI Service gives customers advanced LLM AI with OpenAI GPT-4, GPT-3, Codex, and DALL-E models with the security and enterprise service of Azure. Moreover, Azure OpenAI does not use enterprise data to train the model. Azure OpenAI co-develops the APIs with OpenAI, ensuring compatibility and a smooth transition from one to the other. Microsoft usually has access to new capabilities a few months after release by OpenAI. Data Sovereignty: Currently Azure OpenAI is not available within the Australia region, so use cases would need to be carefully considered for out of region access, given that the inputs are stored in region cache for 30 days as part of their data abuse detection and monitoring facility. **All the other limitations and risks listed here apply equally to Azure OpenAI and open access LLM AI.**	• Continually monitor Azure Open AI security practices. • Create information classification policies for the Azure OpenAI service. • Consider the data sovereignty risks: ○ Create limited use cases for out of region access. • Ensure the general, data integrity and accuracy, legal and privacy, and ethics and fairness risks outlined in this guideline are properly considered.

Service Availability Agencies should identify the availability requirements of LLM AI based on defined use cases and assess the impact to business operations should the service become unavailable.	• Conduct business impact analysis (BIA) on the AI services and document in business continuity plans (BCPs). • Standardise the procedures of using the LLM AI tools and implement alternative methods and workaround for service interruptions. • Where appropriate, use LLM tools as one output to your work (i.e., to supplement and accelerate, not replace).

Dos and Don'ts

SA Government staff must not input any Personal Information or any official government information that is not suitable for public consumption into publicly available AI tools.

DO	DON'T
Assume that inputs into online LLM AI tools will automatically be considered as available in the public domain.	Enter any information (questions or prompts) that shouldn't be in the public domain, including any information that is confidential.
Ensure you consult with your legal team before use to ensure you can comply with any terms of use which may change at any time.	Enter any personal information about any person.
	Enter any health information about any person.
Remain responsible and accountable for your work, including critically assessing all AI generated outputs and validating quality and accuracy with other sources.	Enter inputs or use outputs in a way that may breach another person's intellectual property rights.
Thoroughly research and source reliable references for content where necessary.	Use these tools for any query that is complex or sensitive, or where local context and nuance is critical.
Meet all your existing employment obligations- including in relation to handling official information, privacy, security, human rights, anti-discrimination, administrative and other laws	Ask these tools to answer a question you don't already know the answer to or cannot validate the answer to - you need the knowledge to decide whether it can be trusted or contains bias.
Where required, attribute content that has resulted from the use of these tools	Use these tools to replace your own research, analysis and content development or embed them into your work in a way that means it cannot be done without them.
Regularly monitor and review uses of these tools to ensure they are being used responsibly ethically, safely, and according to law and government policy.	'Copy and paste' sections of AI-generated content into your work. If you do copy and paste any AI generated content, consider what your obligations are in relation to attribution and intellectual property.

Aboriginal Impact Statement

The needs and interests of Aboriginal people have been considered in the development of this guideline. There is no specific impact on Aboriginal people.

Related documents

- Australia's AI Ethics Principles
- SA Government Information Privacy Principles
- Code of Ethics for the SA Public Sector
- South Australian Cyber Security Framework
- PC 030 – Protective Security in the Government of South Australia
- South Australian Protective Security Policy Framework

DOCUMENT CONTROL

Approved by: CIO Steering Committee	
Contact: Office of the Chief Information Officer	Email: CIOAdministrator@sa.gov.au
Review number: v1.0 Final	Compliance: Optional
Next review date: 31 May 2024	Date of approval: 31 May 2023
	Objective Id: B1487266