Interim guidance on government use of public generative AI tools - November 2023

Updated on 22 November 2023

Notice

This guidance will be iterative. It is provided for government agencies to implement within their organisation. APS staff should follow their agency's policies and guidance on using generative AI tools in the first instance.

Feedback from public consultation on the responsible use of AI in Australia will be used to inform consideration across government on appropriate regulatory and policy responses that may include future iterations of this guidance.

Guidance for Australian Public Service (APS) staff

Generative AI tools present new and innovative opportunities for government. However, due to their rapid evolution and uptake, the risks involved in their use need to be considered and assessed.

The breadth of government activities includes developing policy advice for ministers, delivering programs to industry, providing services to the community and providing regulatory oversight. As such, the risk of using generative AI tools for official activities is context-specific and requirements will differ depending on how they are deployed.

Users should first and foremost align with their departmental or agency ICT obligations and policies. The DTA encourages departments and agencies to review their policies related to AI in line with this advice.

This guidance will be supplemented in due course with a risk framework to assist with the risk assessment process.

Golden rules

As you consider using generative AI tools in your work, you should assess the potential benefits and risks for each use case and take appropriate steps to mitigate them.

The principles, tactical guidance and use cases that follow will guide responsible application of these tools. Above all, apply these two **golden rules.**

- You should be able to explain, justify and take ownership of your advice and decisions.
- Assume any information you input into public generative AI tools[1] could become public. Don't input anything that could reveal classified, personal or otherwise sensitive information.

Principles in practice

This section provides guidance to help APS staff adhere to Australia's AI Ethics Principles when using generative AI tools, and is organised into sub-sections as follows:

1. Accountability
2. Transparency and explainability

3. Privacy protection and security
4. Fairness and human-centred values
5. Human, societal and environmental wellbeing

APS staff are encouraged to read and understand Australia's AI Ethics Principles.

1. Accountability

Accountability is one of Australia's AI Ethics Principles and one of the five APS Values. To remain open and accountable to the Australian community, you should be able to explain, justify and take ownership of your advice and decisions.

Generative AI tools must not be the final decision-maker on government advice or services.

You may consider using generative AI tools to brainstorm options, generate code or draft content. However, these tools can produce a convincing but inaccurate response to a query, sometimes referred to as 'hallucination'. A human with the appropriate skills, knowledge or experience should review output before using it (especially before entering any coding outputs into government systems).

Finally, you should undertake appropriate training to enable you to critically analyse the outputs produced by generative AI tools and to understand the limitations of the technology.

2. Transparency and explainability

It should be clear when generative AI tools are being used by government to inform activities. Users could consider including markings in briefings and official communications indicating if generative AI was used to generate any of the information.

Building on the accountability principle, you should critically examine outputs from generative AI tools to ensure your advice and decisions reflect consideration of all relevant information and do not incorporate irrelevant or inaccurate information.

Remember, you should be able to explain, justify and take ownership of your advice and decisions.

3. Privacy protection and security

Inputs into public generative AI tools should not include or reveal classified, personal or otherwise sensitive information. All activities need to align with legislation and policies relating to information and data (for example the *Privacy Act 1988*, and the Protective Security Policy Framework).

Government information must only be entered into public generative AI tools if it has already been made public or would be acceptable to be made public. Employees determining that the information in question is suitable for public release must have the appropriate organisational delegation to do so.

Classified or sensitive information must not be entered into these tools under any circumstances.

You should not enter information that would allow public generative AI tools to extrapolate classified or sensitive information based on the aggregation of content you have entered over time.

Any data entered into public generative AI tools is stored externally to government and we do not know who has access to it. Where available, you should disable any settings or permissions which save chat history.

4. Fairness and human-centred values

Generative AI tools are typically trained on broad sets of data that may contain bias. Bias can arise in data where it is incomplete, unrepresentative or reflects societal prejudices.

Generative AI tools may reproduce biases present in the training data, which could lead to misleading or unfair outputs. Bias in outputs may disproportionally impact some groups, such as First Nations people, people with disability, LGBTIQ+ communities and multicultural communities.

Before using AI-generated outputs, you should consider whether you have a process in place to ensure that outcomes are fair and meet community expectations. For example, you could include representatives of relevant communities (with appropriate skills, knowledge or experience) in decision-making on use of data and outputs relating to those communities.

Remember, you should be able to explain, justify and take ownership of your advice and decisions.

5. Human, societal and environmental wellbeing

You should engage with generative AI tools to better understand the potential benefits and risks of the technology, and the role it can play in contributing to the strategic priorities of your agency and the Australian Government.

You should weigh the benefits and risks of a particular use of generative AI, and consider whether it is the right tool for the job. Traditional tools may be cheaper, safer or better suited to the task at hand.

You should only use generative AI tools in a manner consistent with the APS Values, Employment Principles and Code of Conduct for purposes that are consistent with improving the wellbeing of the Australian community. You should also consider intellectual property rights of third parties as well as broader copyright issues when using these tools and seek legal advice where necessary.[2]

Consider whether your use of Indigenous data and generative AI outputs is consistent with the expectations of First Nations peoples, particularly around Indigenous data sovereignty and governance, and with the forthcoming Framework for Governance of Indigenous Data.

Tactical guidance - dos and don'ts

- Check whether your use of generative AI aligns with your departmental or agency ICT obligations and policies, including whether you are required to obtain approval or register your use of the platform before you use it.
- When you are using a public generative AI tool that requires a login as part of your work, you should use your work email to sign up or log in, and create a unique password (do not use your work account password). If a public generative AI tool can be used without needing to create an account, then don’t create an account.
- Do not distribute or click on any links provided or generated by public generative AI tools or bots. These links could lead to phishing sites or malware downloads. Only click on links from trusted sources.
- Treat with care any files generated by public generative AI tools or bots which have the potential to contain malicious code, such as macros in Microsoft Office files. These files must be considered as potentially malicious, and not interacted with or distributed to other people, until vetted and proven to be non-malicious.
- Where developing tools or processes incorporating generative AI (for example, to summarise information for a daily news update), be sure to continue to monitor performance against intended purpose. Generative AI models are updated from time to time and may learn from the data they interact with. This may affect performance and reliability of output.
- Align to relevant principles, guidelines and best practice issued by the Australian Government, including:
    - Australia's AI Ethics Principles - Department of Industry, Science and Resources
    - Artificial Intelligence guidelines - Australian Government Architecture
- You should report any instances where you are not able to fully apply this guidance to your agency's relevant senior accountable officer (e.g. Chief Information Security Officer or Chief Information Officer).

Use cases

Generating 'first pass' content

Nick needs to develop a project plan and is wondering if he can use ChatGPT to create a baseline project plan that he can then improve upon.

What should Nick do?

Nick can get the template of a project plan from ChatGPT. Nick must refrain from entering any details of the project, such as the project name, agency, names of systems/software, high level requirements or staff members involved. These details could provide sensitive information about the project to ChatGPT.

Generating files or documents

Stephanie is creating an urgent report to present later in the day and realises she is short of time to make a PowerPoint presentation. She is planning to upload the data for the slides to a public AI platform to help create a quick presentation.

What should Stephanie do?

Stephanie must not upload or input any sensitive or classified information into public AI platforms. She could input non-sensitive information, such as information available on a government website or an agency report that has been released publicly. She could also ask the AI for generic slide templates to help her prepare her presentation.

Using generative AI-powered search

Roque is writing technical requirements for a tender that needs to go out urgently. Roque wants to type his requirements into Google Bard to confirm some of the technical specifications around monitor resolutions.

What should Roque do?

Roque can use Google Bard to find information about technical aspects, however Roque should take care not to input any details of his agency's specific requirements or any organisational information. Roque should also take care not to mention the word tender or any market sensitive information that could, combined with his official email address, indicate a tender is being prepared by his agency. The information received should also be thoroughly validated by a human for appropriateness and accuracy before being used.

Exploring datasets

Angus is doing some basic data analysis on a publicly-available dataset. He is wondering whether he can use ChatGPT Plus to help him generate some insights from the data.

What should Angus do?

Given the data Angus is using is publicly available and there is no particular sensitivity to the topic he is researching, Angus can use ChatGPT Plus to assist him with data analysis.

However, Angus should remember that generative AI tools can produce output that is inaccurate or biased. Before sharing more broadly or making any use of the insights, Angus should check that the output fairly and accurately reflects the data.

Footnotes

[1]Public generative AI tools are widely available third party AI platforms, tools or software (whether accessed through a browser or through an application), such as ChatGPT and MidJourney, which have not been security risk assessed by your agency and approved for use with classified or sensitive data. This also includes the enterprise version of Microsoft Copilot (previously referred to as Bing Chat Enterprise), where agencies have enabled it for staff use.

[2]Note that some AI service providers have announced that they will provide indemnity for IP liability in some circumstances. See e.g. Adobe Firefly (https://www.adobe.com/au/sensei/generative-ai/firefly/enterprise.html) and Microsoft (https://blogs.microsoft.com/on-the-issues/2023/09/07/copilot-copyright-commitment-ai-legal-concerns/).