Generative-AI: guidelines for researchers

The recent and rapid emergence of Generative-AI (Gen-AI) tools presents many opportunities for research. The benefit these tools may bring to research and contribute to the public good is being explored by researchers across all disciplines. While 'the sky is the limit' in what these new tools may offer, there are identified risks that must be managed so that Gen-AI tools can be used safely in research.

These guidelines provide a simple set of considerations that all researchers need to recognise before using Gen-AI tools in research.

What is Generative-AI?

Gen-AI refers to a set of algorithms, data and applications that analyse and generate text, images, code and other outputs based on models created from immense amounts of content available on the internet (e.g. Wikipedia, open datasets, websites, social media platforms, image banks). Well-known Gen-AI tools include Large Language Models (LLMs) such as ChatGPT, Copilot (formerly Bing Chat), Bard, and Llama; and image generators such as Midjourney, Dall-e, Stable Diffusion and Adobe Firefly, but there are thousands more available now and under development.

What are the top risks to consider when using Generative-AI in research?

1. **Handling Sensitive Information:** Uploading (or entering or copying) unpublished research content into openly available cloud-based Gen-AI apps may unintentionally make that content **'open access'**. Once information is fed into a Gen-AI tool, it could be incorporated into the training datasets that the tool uses to generate content for all users and could therefore be presented to another user at a later date. The risks are:
    - loss of Intellectual Property (IP),
    - breaches of confidential or sensitive information (including privacy breaches),
    - non-compliance with regulations, research codes, funding rules, ethics approvals, university and journal policies, and state and federal law.

    It is strongly recommended that researchers use the University's Protected MS Copilot service instead of using openly available tools found online. This service reduces the risk of privacy breaches and loss of IP, and provides a safer environment for research use.

2. **Accuracy:** The content created by a Gen-AI tool can sometimes be inaccurate. It is important to independently verify the information gained from a large language model, just as it is important to check the source's trustworthiness when you do an internet search. The example of a lawyer citing non-existent case law shows what can happen if you don't independently verify your sources. Model responses can also be

biased towards more prevalent culture and opinion in the training data, for example in the original version of ChatGPT, about 93% of the training data is in English.

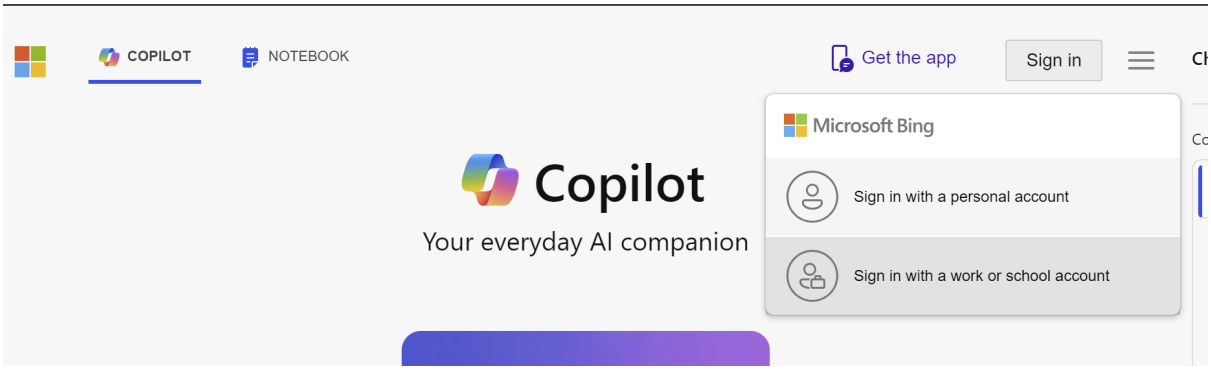
3. **Accountability** The accuracy and truthfulness in research – be that writing, data analysis or decision making – lies with the researcher using the AI and not with the AI itself. The researcher must ensure they are not lying, fabricating, misleading or misrepresenting their findings, and they cannot delegate this responsibility to an AI. A researcher using Gen-AI needs to understand the proper way to attribute such work based upon the policy of their institution, grant funding body, journal and relevant government agencies.

How can I use Generative-AI safely?

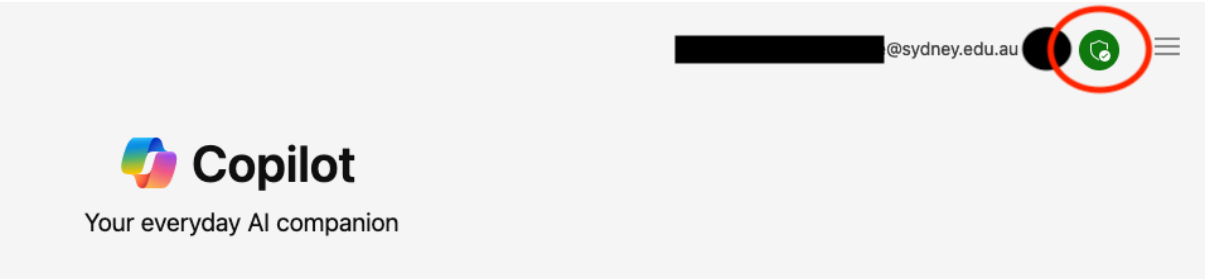
1. **Use the University's Protected MS Copilot service**

The University now hosts a more secure Gen-AI service, Microsoft Copilot (for Web) https://copilot.microsoft.com, which has been assessed as being suitable for research data that has a 'Protected' classification. All researchers are encouraged strongly to use this service instead of the 'open' Gen-AI tools (e.g ChatGPT).

It is important that researchers login to the Protected Copilot service ***using their university email address and select the 'Work account' option to trigger the University's secure unikey and OKTA login***.



Once logged in, the green 'Protected' sign (a green circle with a shield and a tick) will appear in the top right of the page, as shown here:



If the 'Protected' sign is absent then the Protected Copilot service has not been accessed. The researcher should logout and login again using the above instructions.

Note that sensitive and confidential information, including ***all research data that is classified as 'Highly Protected'***, ***must not be uploaded*** (copied into the Chat prompt) into Copilot.

2. **Using openly available Generative AI tools (e.g. ChatGPT)**

- **Ensure that the only content you are uploading is material that can be shared safely with anyone.**
    - **Only upload data that has 'Public' classification into an openly available Gen-AI tool (e.g. ChatGPT)**
    - Do not upload unpublished research findings into a cloud-based Gen-AI tool. For example, using ChatGPT or Elicit to assist in writing a literature review, to help brainstorm or simplify ideas is probably low risk, but writing the final text of a discussion, synthesis of research findings or entire thesis is high risk.
    - Do not upload research data that contains confidential or sensitive information into a cloud-based Gen-AI tool (this includes personally identifying information of study participants, commercially sensitive information, and culturally or environmentally sensitive information).
    - Do not upload your personal data into a cloud-based Gen-AI tool.
    - In online services like ChatGPT, make sure to opt-out of having your data used for training future models.

Guidance for using ALL Generative AI tools

- **Validate that the output from Gen-AI tools is accurate and attributable.**

  Gen-AI tools use probabilistic models that are trained on enormous datasets to generate plausible new content; they use existing, uncorroborated content from many sources and can produce inaccurate, biased or creatively fictitious content. Repeated use of a Gen-AI tool can create different content each time it is run. If using Gen-AI tools to generate research content, researchers need to ensure that outputs are accurate, factual, fair, able to be attributed properly according to institutional, publisher and funding body policies, and that research results are replicable.

- **Ensure that you can demonstrate your contribution, and that of the AI, to all research processes and outputs**

  As a researcher, you must be able to verify and replicate your research processes and findings. It is important that you track how you have used a Gen-AI tool in your research. **It is strongly recommended that you take screenshots of all inputs that you make into Gen-AI Chat prompts, and keep records of all generated outputs**. You may be required to quantify or otherwise describe the contribution that Gen-AI has made to your research.

- **Ensure that you are allowed to use Gen-AI in your research.**

  Before using a Gen-AI tool for research, check that your ethics protocols, funding agreements, data sharing agreements or commercial contracts permit use of Gen-AI in particular, or more generally, cloud-based services in other jurisdictions.

- **Check the Gen-AI policy of the journal that you plan to publish in.**

  Some journals prohibit use of Gen-AI, and many have specific content rules. For example, Nature Publishing Group allows use of Large Language models in published work but disallows use of generative images or video.

- **Observe confidentiality when using Gen-AI**

  Protect the IP of researchers and maintain professional standards: ***do not use Gen-AI when peer-reviewing grant applications, manuscripts and publications***.

- **Ensure your students are aware.**

  If you are supervising a research student (HDR or Honours), make sure that they are aware of the risks of using Gen-AI in their research, including proposal and thesis writing.

More information

For expanded guidelines, see this Knowledgebase article

Queries? Contact the Research Service desk