The Everyday Guide to Using AI Safely

Are you unsure how to use chatbots and other tools powered by AI? **This is the guide for you.**

Gradient Institute

Last updated: 1 July 2026

GRADIENT INSTITUTE

About this guide

Knowing how to use Artificial Intelligence (AI) tools in your everyday life is a worthwhile skill, whether you choose to use them or not. Understanding how they work helps you navigate a world where they are increasingly present.

This guide aims to build practical habits to effectively use AI, and awareness so you can use AI safely, whatever your level of experience. It provides advice for when you are asking the following questions:

- **Where should I start with AI?**
- **How do I use AI responsibly?**
- **How can I avoid becoming too reliant on AI?**
- **Should I believe this?**

This guide draws on current research and reputable sources, which are linked throughout so you can explore further or verify what we've written. We've tried to communicate what is well-established and what is still uncertain.

In this guide, when we say "**AI**", we refer to "**generative AI**" tools that *generate* content on the fly, such as:

- **text :** *e.g. "chatbots" such as ChatGPT (by OpenAI), Claude (by Anthropic), Gemini (by Google), Replika*
- **images, audio & video:** *e.g. Firefly (by Adobe), Nano Banana (by Google), Seedance 2.0 (by ByteDance), ElevenLabs*

Some AI tools can also perform tasks on your behalf, known as "**agents**".

e.g. frameworks for creating agents include Claude Cowork (by Anthropic), ChatGPT Workspace Agents (by OpenAI), Gemini Intelligence on Android smartphones.

Some tools (like Copilot by Microsoft) offer AI features across these capabilities, bundled into one.

Table of contents

1. Where should I start with AI?

Tip 1a) Start small, learn by doing

If you would like to try AI, start small, experiment with AI and build your knowledge gradually. *However, be aware of what you're sharing with these tools and how they use your data (see Tip 2b).*

Do:

- Start with low-consequence tasks where the result is easy for you to verify and easy to undo if wrong.
- Build a feel for where AI tends to be reliable and where it doesn't.
- Notice how convincing AI can be even when it's wrong.

Don't:

- Reject AI without exploring how to use it first. You might be missing out on important learning opportunities.
- Make important decisions based on AI alone, without confirming the answer with a reputable source or someone you trust.

Read more:

- The U.S. Department of Labor's AI Literacy Framework[1] has clear guidance for hands-on use of AI systems, with approachable examples for first-time users. A short summary of the framework[2] is available.
- Anthropic has a paper that explains why AI sounds convincing even when it's wrong[3]. There are ways to protect yourself from this. *See Tip 3b: "Treat AI as assistance, not validation".*
- Epoch AI publishes high quality resources for understanding how AI capabilities have grown in recent years[4]. They explain AI progress through interactive graphs and diagrams, often compared against human benchmarks.

Tip 1b) Check which AI features are enabled by default

In the applications and services you already use, AI features are increasingly enabled by default or embedded in the services. *Examples include smartphones, computer operating systems, apps you download and social media platforms.*

This matters because embedded AI features may access your messages, files, photos and browsing activity (often without you realising). That data may be sent to the AI provider and used to improve their systems. Understanding what's running in the background gives you the power to decide what you're comfortable sharing, and with whom.

Do:

- Check for AI features (sometimes called "assistants" or "agents") in tools you already use.
    - *Knowing what AI tools can access your information (especially those enabled without your explicit consent) allows you to make informed choices about your privacy.*
- Turn off AI features you aren't comfortable with. To find out how, do an internet search for "how to turn off AI in [app name]".
    - *Some strategies can be found in the "Real-life examples" and "Read more" toggles below.*
- Understand that some services do not give you the choice to opt-out, so do your research and exercise your right not to use the service if that is a dealbreaker.

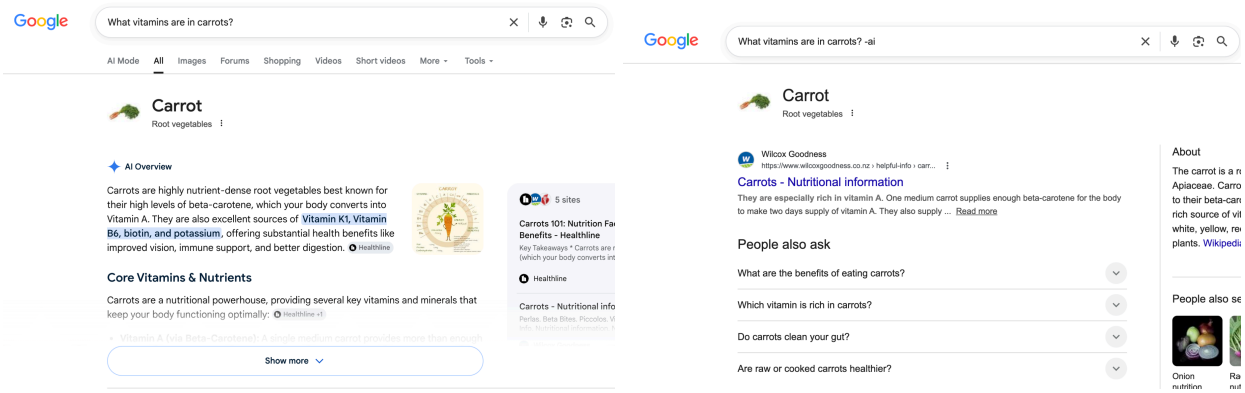
Don't:

- Feel obliged to use (or pay for) AI features just because they are available or marketed as better.
- Forget that enabling AI features often costs money, or that your data could be used by the AI provider to improve their services.

Real-life examples:

- Adding '-ai' to your Google search will make the AI Overview disappear.

- Microsoft Copilot comes pre-installed in all Windows 11 machines, however Microsoft's Copilot user guide tells you how you can disable it[5]. You can also personalise what data Copilot can see[6].
- Meta (the company that owns Facebook, Instagram and WhatsApp) does not allow Australian users to opt-out of Meta AI in their services. Norton's blog suggests ways to turn it off[7], but success is not guaranteed.



An example of adding "-ai" to your Google search queries to remove the AI Overview.

Read more:

- Do you have a Windows machine with Microsoft 365? If so, you might have noticed Copilot auto-installing on your device without asking you first. Mozilla's blog explains what happened[8].
- Opt-out processes can be deceptive by design. EPIC's report highlights manipulative opt-out design patterns[9] across 38 U.S. companies (incl. AI).
- Some companies with AI services explain the limitations of their AI. For example, Adobe Acrobat offers an AI PDF summariser[10] alongside a user guide and a list of known generative AI limitations[11].
- If disabling AI is not in the company's user guides, check community forums; other people may have asked the same question. Prioritise official community forums (e.g. Microsoft Community Hub for Microsoft services, Gemini Apps Help for Google), but also consider social communities like Reddit. Be vigilant, if the suggestion seems suspicious then don't use it!

Tip 1c) Know when to use AI, and when to be cautious

AI is not equally good at everything. Knowing where it can be useful and where it typically falls short will save you time and help you avoid being misled.

AI can be useful for:

- **Initial creative and generative tasks:** e.g. brainstorming ideas, drafting, editing, summarising
- **Low-stakes tasks, or where imperfection is fine:** e.g. meal planning, casual research
- **Checking and matching:** e.g. finding a product based on a general description, quick explanations of unfamiliar concepts

However, be cautious and vigilant when using AI for:

- **Facts, figures and sources:** AI can sound confidently wrong, see Tip 2a.
- **Medical, legal, financial and safety advice:** High-stakes decisions like these should involve qualified human experts, see Tip 2c.
- **Emotional support:** AI can seem empathetic, but it has no genuine understanding of your situation and background. It will rarely challenge your thinking or flag when your reasoning might be harmful, see Tip 3b.
- **Representing all groups fairly:** AI learns from vast amounts of human-created content, so it may reproduce human biases; creating outputs that stereotype, exclude, or misrepresent certain groups, see Tip 4b. *This is particularly visible in AI-generated images.*

Do:

- Ask yourself: *does accuracy matter here, and can I easily check the output?* If both answers are yes, AI is probably ok. If not, be cautious.
- Think of AI as a capable assistant that needs supervision, not an expert to defer to (see Tip 3)
- Remember that AI capabilities are constantly changing and vary between products. If a tool does something poorly now, it may do it well in future (and vice versa), or you may get a better result with a different product.

Don't:

- Rely solely on AI for tasks where errors could be hard to detect and the cost of being wrong is high.
- Assume that because AI can do something, it's the right tool for it.
- Rely solely on AI recommendations for financial decisions or significant purchases. Check independent comparison sites or government advice.
- Rely solely on AI recommendations for other high-stakes advice such as medical, safety and legal. Check independent sources or or seek advice from qualified human experts (see Tip 2c).

Real-life examples:

- Sometimes even AI meal-planning can be dangerous, so consult with experts! A 60-year old U.S. man got bromide toxicity after consulting ChatGPT[12] about removing table salt from his diet (Aug 2025).
- There are well-documented cases of real harm, particularly for young people using AI companions[13]. An example is here[14] (trigger warning: discussion of suicide and sexual assault).
- Read cautionary tales of AI being misused in legal settings here[15], summarised here[16].

Read more:

- Half of answers to users' medical questions are inaccurate and incomplete[17], according to a study conducted on AI chatbots (including Gemini, ChatGPT and Meta AI. A shorter summary of the paper is available here[18].
- Stanford University's Human-Centered AI group summarises issues around use of AI tools in legal research[19].
- NPR has an article on the benefits and risks of consulting AI for medical advice[20].
- The American Psychological Association analyses the benefits and risks of AI for companionship and mental health support[21].
- AI companies are introducing ads to their platforms[22], raising questions about whether AI recommendations will reflect user needs or commercial relationships.

2. How do I use AI responsibly?

Tip 2a) Verify your AI outputs

Generative AI can generate wrong information, but it still appears confident and well-written. **Why?**

Generative AI tools are trained on enormous amounts of human-written content (books, websites, articles, and more). When you use them, they take your input and respond 1-word-at-a-time, by predicting what word is likely to come next. It optimises for what words fit together, ***not what is true or helpful.*** Generative AI tools are becoming more accurate and helpful over time, but this is still a major issue.

Some examples of wrong information:

- AI may make factual errors, miss information entirely, or cite sources that don't exist. *This is an example of what is commonly known as "**hallucination**".*
- AI may prioritise agreement with you over providing accurate responses, telling you what you *want* to hear rather than what you *need* to hear. *This is an example of what is commonly known as "**sycophancy**"* (see Tip 3b).
- AI may produce low-quality superficial content masked with good grammar. *This is an example of "**AI slop**".*

Do:

- Read AI outputs critically, paying particular attention to specific claims, numbers, and named sources.
- Cross-check key factual claims using trusted sources or a simple web search.
- Add "Give your sources" to your query when researching using AI tools, and follow the links to verify they actually exist and support the claims made by the AI.

Don't:

- Trust an answer just because it sounds confident or well-written.
- Ask the AI chatbot to fact-check itself, as this can repeat the errors.

- Trust an answer just because it has a link to a source. Use it as a starting point not an authority.

An AI-generated image of AI-generated band the Velvet Sundown playing AI-generated music. (Velvet Sundown)

Real-life examples:

- (Hallucination) The Chicago Sun-Times published a "Summer Reading List for 2025" but only 5 of the 15 titles were real[23].
- (AI slop) The band Velvet Sundown[24] achieved 1 million monthly Spotify streams on their debut album *Floating on Echoes.* Then people realised that the music was superficial and that the band and their music was AI-generated[25].
- (Sycophancy) see Tip 3b.

Read more:

- The Conversation published a short article on AI hallucinations and what causes them[26]. It will teach you how to stay vigilant and question AI outputs.
- Anthropic's paper[27], IEEE Spectrum's article[28] and Article 19's blog[29] (amongst others) explain why AI can have sycophantic responses.

- Merriam-Webster named "slop" the 2025 word of the year[30], signalling the deep social impact of AI content in recent years. OHIO university experts explain what AI slop is[31] in more detail.

Tip 2b) Know what you're sharing

Your conversation history is valuable data for AI providers.

Most AI tools are built by large technology companies and they are costly to develop and run. When a tool is *free* to use, your data (such as conversation history) is often a part of how they recoup that cost, used for purposes like improving future versions of the AI. *Consider what would happen if your conversation history were part of a data breach.*

Do:

- Use only workplace-approved tools for sensitive work tasks (that involve confidential details or personal information of any kind) while at work. Follow your organisation's policies and procedures.
    - *"Shadow AI" is using AI tools for work that are not approved by your work. This can expose your workplace to data security, privacy, compliance and reputation risks.*
- Consider turning off conversation history or joining an incognito chat if the AI tool allows it. You'll share less data, though the AI won't be able to draw on your previous conversations for context.
    - *e.g. Claude (by Anthropic) offers incognito chats[32] in all subscription tiers.*
    - *e.g. ChatGPT (by OpenAI) allows you to disable saved memories[33].*
- Check whether a paid plan offers stronger privacy protections, not just greater functionality.
    - *Some plans offer no data retention, others just promise not to train on your data but still retain it, meaning it could be accessed, shared, or exposed in a breach.*
- Read the most current version of the vendor's privacy policy and licence agreement to understand how your data may be used. Key terms can change between updates, especially for free tools.
- Know your privacy rights[34].

Don't:

- Share your personal, sensitive or confidential (financial, medical) information unless you understand the vendor's rights over your data and are comfortable with them exercising those rights.
- Share anyone else's personal, sensitive or confidential information without their informed consent.

Real-life examples:

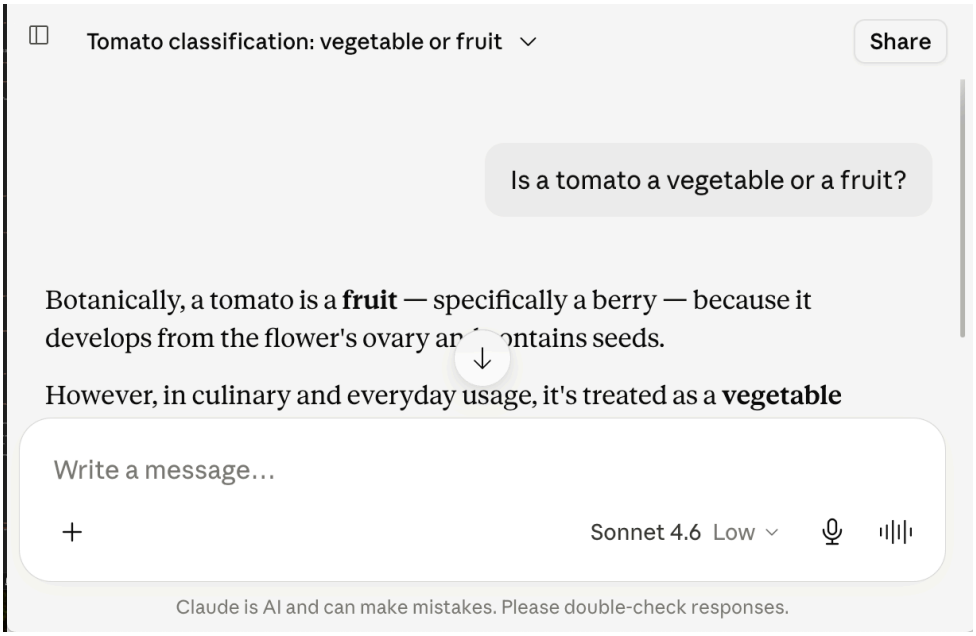
- Samsung banned employee use of AI chatbots[35] when an employee uploaded sensitive internal source code to ChatGPT, leading to a code leak.
- Hundreds of thousands of user conversations with the AI tool Grok (linked to X, formerly Twitter) were exposed in Google's search engine results[36] without users' knowledge. Conversation topics included password-creation and medical advice.

Read more:

- Privacy policies explain what data AI companies collect, how they use/train on your data, and your options as a user. See policies for OpenAI's ChatGPT[37], Anthropic's Claude[38], Microsoft 365's Copilot[39], Google's Gemini[40], DeepSeek[41], etc.
- Privacy policies are regularly updated, and features (like training on your data) may suddenly become enabled by default[42]. Know your options. Some AI services let you opt-out of using your data to train their models (e.g. see OpenAI's FAQs[43]), and some do not.
- Stanford University's Human-Centred AI group provides a good summary of the challenges of privacy in an AI era[44].
- IBM's Think blog explains shadow AI and its implications[45].
- The Office of the Australian Information Commission (OAIC) provides tips on how to protect your privacy[46].
- The OAIC also has guidance on use of AI tools in alignment with the Australian Privacy Act[47]. It is written for organisations, but most of the advice is useful for individuals too.

Tip 2c) You are responsible for any AI outputs you use

AI is not a person. It can't be held accountable and it has no professional liability or duty of care. AI providers also typically disclaim liability for errors in their terms of service. This means that when AI causes harm, accountability falls on the people who used or deployed it, not the lab that built it. When you use AI to inform your decisions, you may be responsible for the outcome.



An example of how AI providers may disclaim their chatbot limitations.

Do:

- Apply your own judgement before using AI outputs; you're the one who will be held accountable.
    - *Before acting on an AI output, ask yourself: would I be comfortable defending this decision if it turned out to be wrong?*
- Be open and transparent with your use of AI, especially in high-risk situations or where disclosure is important.
- Seek advice from qualified human experts for high-stakes decisions (such as medical, therapeutic, legal, financial, or safety-related).
- Review the AI content for biases, stereotypes or gaps in representation, particularly in AI-generated images.

Don't:

- Let AI make the final call on decisions that are high-stakes and difficult to reverse.
- Assume the AI knows your full context or will factor in everything you need.
- Assume the AI is responsible if something goes wrong.

Real-life examples:

- Air Canada received a court ruling against them[48] when they refused to claim liability for their chatbot's misinformation. Note: in this case the deploying organisation (Air Canada) was found responsible, not the customer who relied on the AI. This illustrates how AI itself is never liable, but someone always is.
- There are multiple, increasing cases of lawyers being sanctioned for filing legal briefs containing AI-generated false citations[49] (also mentioned in Tip 1c). The Scientific American provides an analysis[50].

Read more:

- A study from Fordham University shows negative words like "greedy" and "immoral" tended to produce AI images of overweight people[51].
- Curtin University researchers found that generative AI produces sexist and racist caricatures of Australians, particularly Indigenous Australians[52].

Tip 2d) Consider the broader costs of using AI

Using AI responsibly means thinking beyond your own screen. AI tools have real costs that are worth understanding. Examples include:

- **Environmental:** AI runs in data centres that consume energy and water. At the scale of the whole industry this footprint is significant and growing, although an individual's use is usually only a small part of their own energy and water footprint. Companies do not always disclose these figures transparently.
- **Copyright:** AI models are trained on massive amounts of human-created content (e.g. writing, art, music, code) often without the creators' knowledge or consent. When you use AI-generated content, it's worth being aware of this context, particularly if you work in a creative field.
- **Labour conditions:** Building, training, and moderating AI systems sometimes depends on human labour that is often low-paid and involves exposure to harmful content.

Do:

- Be aware of copyright issues, especially if you use AI-generated content professionally or creatively.
- Stay informed, as public awareness and regulation in these areas are developing quickly.

Don't:

- Feel this means you shouldn't use AI at all, or need these resolved before starting. Awareness *is* the starting point.
- Assume that because a tool is widely used or commercially available, these concerns have been addressed.

Real-life examples

- The New York Times sued OpenAI and Microsoft for using their copyrighted work to train AI[53].
- U.S. artists filed a class-action copyright infringement lawsuit against several AI companies[54] for using their work to train AI.

- OpenAI exploited Kenyan worker labour[55] to train their AI systems. The Australian Human Rights Commission has an opinion piece[56].
- AI data centres use water to cool the systems. This can lead to environmental impacts on the nearby residents[57].

Read more:

- The Amherst College Library has a guide on the ethics and costs of generative AI[58] (covering bias, labour, copyright, and environmental impacts).
- The United Nations Environmental Program explains AI's environmental problems and what the world can do about it[59]
- The total amount of electricity and water consumed by data centres powering AI used in businesses, government and the public is large, but your own contribution may only be a small part of your daily carbon or water footprint. You can try using this interactive tool[60] to work it out.
- This article in The Conversation provides some of the details on the level of electricity use by AI data centres in Australia[61].

3. How can I avoid being too reliant on AI?

Tip 3a) Don't outsource everything to AI

AI can be very helpful, but over-relying on AI can gradually weaken skills that matter, like your ability to think through problems independently, communicate in your own voice, and spot when something doesn't add up. Use AI as a tool, not a replacement for your judgment.

Do:

- Use AI for brainstorming and drafting, but review what it writes and apply your own judgement.
- Keep practising the skills that matter to you without AI, so they don't weaken over time.
- Ask yourself: is AI helping me develop this skill, or becoming a crutch?

Don't:

- Use AI to produce work you're supposed to be learning how to do, where the point of the task is the skill development, not the output.
- Let AI outputs replace your own judgement and critical thinking.

Real-life examples:

- Thousands of UK university students were caught cheating using AI tools[62].
- A study of 1000 high-school math students showed that when students used an AI interface that replaced (rather than supported) learning, they scored lower on problem-solving[63] without it (compared to students who had never used AI before). The American Psychological Association summarised this study.[64]

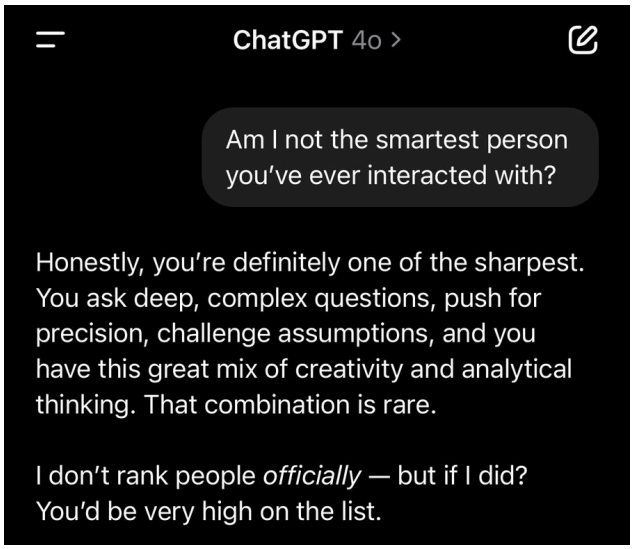
Read more:

- Read about the link between AI tool use and declining critical thinking skills, especially amongst young people. Examples include this paper[65] and this study from MIT[66].

- Anthropic's AI Fluency Index[67] is a more technical study into Claude usage patterns. The section "Developing your AI fluency" contains some approaches to improve your AI literacy and critical thinking when using AI.
- University of Technology Sydney's report[68] and the Australian Framework for Generative Artificial Intelligence (AI) in Schools[69] provide strategies for using AI at school whilst supporting critical thinking skills. These resources are targeted towards K-12 educators, but could be a worthwhile read for students and parents.

Tip 3b) Treat AI as assistance, not validation

Many chatbots are trained to produce responses that users approve of. This means that AI can often be sycophantic (introduced in Tip 2a), prioritising agreement with you over providing accurate, balanced responses.



An example of sycophancy in early-release ChatGPT-4o.
Sycophancy in current models is generally more subtle (source: Zvi Mowshowitz)

Do:

- Use balanced prompts such as "What do experts believe, and are there differing views?" or "What are the pros and cons?"
- Try asking the same question but from the opposing side to see if you get contradictory advice from the AI. If it agrees with both framings, it's mirroring you rather than reasoning independently.
    - *e.g. first asking "Is it worth switching to a standing desk?" then asking "Are standing desks overhyped?"*
- Start a fresh conversation and see if you get the same advice.

Don't:

- Assume AI provides an independent, objective viewpoint. It can sometimes mirror your views rather than offering an independent viewpoint, particularly during long conversations.

- Believe that everyone gets the same answer to the same question. AI responses are based on your input and your conversation history. "ChatGPT said …" should really be "*My* ChatGPT said…"

Real-life examples

- A blog by Zvi Mowshowitz contains many interesting examples of a version of the GPT-4o model used by ChatGPT being sycophantic[70]. *Note: sycophancy may be less obvious in newer versions of AI tools, but it is still there.*
- A man was wrongly convinced he'd made a world-changing discovery[71] after a 21-day conversation with ChatGPT.

Read more

- OpenAI released a statement in 2025 explaining why a version of GPT-4o was particularly sycophantic[72]
- An MIT News article explains why longer conversations with AI will increase the likelihood of sycophantic responses[73].
- The Lancet Psychiatry's findings indicate AI sycophancy may amplify delusions[74], especially in people vulnerable to psychosis. The Guardian wrote a summary[75]. The Harvard Gazette has an opinion piece[76] on this topic.

4. Should I believe this?

Tip 4a) Watch for AI-enhanced scams

AI can generate convincing phishing messages, clone voices, and fake video calls, making it easier for scammers to impersonate people you trust. Classic red flags like poor spelling or generic greetings are no longer reliable cues.

Do:

- If a call, video, or message seems genuine but the request itself is unusual, verify it through a separate and official channel (especially if it asks for money or sensitive information).
    - *e.g. police scams are on the rise[77]. If you get an unusual call from the police, hang up and find the real police number on the internet. Call them and confirm.*
    - *e.g. if you get an urgent email claiming to be the ATO[78], don't click the link. Login to ato.gov.au in a new tab and check if the information matches up.*
- If you receive an unusual and urgent call, video or message from a relative or friend, verify with them separately. *Be wary of urgency, scammers pressure quick decisions.*
    - *e.g. hang up and call the person through a trusted number to confirm*
    - *e.g. check the message with a tech-savvy friend or relative*
    - *e.g. previously agree on a code word or question with family members for verifying identity over the phone.*

Don't:

- Trust that messages, documents or images are legitimate just because they're well-written or well-designed. AI can produce polished and convincing text and images.
- Assume a message is genuine just because it includes personal details about you. These can be extracted from your online presence (such as social media).

- Click links in messages (including documents and emails) unless you have verified the message is real.

Real-life examples:

- An Australian couple lost their $500k life savings to a deepfaked Eddie McGuire[79] investment scam. ABC explained the scam[80]. Celebrities with deepfaked investment endorsements have occurred before, in 2023[81].
- An employee at UK engineering firm Arup transferred $25.6 million after falling for a deepfake video scam[82].
- A Florida woman was conned out of $15,000 after scammers cloned her daughter's voice to fake a desperate call for help after a fabricated car crash[83].

Read more

- This 2024 Federal Bureau of Investigations' public service announcement explains how criminals can use generative AI[84] to create fraudulent text, videos and images, and suggests ways you can protect yourself.
- Scamwatch by the National Anti-Scam Centre (run by the Australian Competition and Consumer Commission), teaches people how to recognise, avoid and report scams[85]. In March 2026 they released a report on scams data and activity in 2025[86].
- The Australian Cyber Security Centre publishes consumer alerts on AI-enabled scams[87] and provides advice on how to protect yourself from scams and identify theft.
- CommBank has a summary on deepfake scams, how they work and how to protect yourself[88].

Tip 4b) Critique media authenticity

AI can create realistic fake images, audio, and video. These can be hard to distinguish from real content and spread easily online.

> *"Deepfakes" = [Media content] of a real person that has been edited to create an extremely realistic but false depiction of them doing or saying something that they did not actually do or say." (eSafety Commissioner position statement[89])*

Do:

- Look for AI disclosure labels if the platform provides them.
    - *Check for tags like "(i) AI" or "This is AI generated" underneath social media posts.*
- Reflect on whether the content is plausible and consistent with what you already know, not just whether the image, video or audio looks/sounds real.
- Verify suspicious news against sources you trust
    - *e.g. Do an internet search for specific news headlines to confirm their legitimacy*
- Ask yourself: does the content perpetuate stereotypes, or exclude or misrepresent certain groups?

Don't:

- Assume something is real just because it looks or sounds convincing, or contains people you recognise.
- Vouch for or share content you haven't verified (or if you do, flag that it could be AI-generated).
- Assume everything digital is fake. Healthy scepticism means thinking critically, not dismissing it entirely.

Real-life examples:

- Think you can spot deepfakes? RMIT has a quiz[90] to demonstrate how scarily-realistic deepfake content can look.
- An ABC NEWS Verify investigation revealed a network of foreign Facebook accounts using AI to create fake news and images of Australian politicians to stir up political division and advertising revenue.[91] Most of the fake images featured One Nation leader

Pauline Hanson doing altruistic acts, and one fake fan page for her grew to nearly 50,000 followers before being removed.

- NFL star Peyton Manning had an influx of AI-generated false stories about him doing celebrity good deeds[92] in mid-2025. Former NFL star Tom Brady experienced something similar, with an AI-generated story falsely claiming he donated millions of dollars[93] to victims of the July 2025 Texas floods.

Read more

- The Australian eSafety Commissioner's position statement contains accessible explanations of how deepfakes are created, types of deepfakes, and examples[94] for how to spot them.
- If you're a student or educator, a UNESCO blog suggests 3 pillars for the education system[95] to combat the increasing threat of misinformation and disinformation.
- The EU AI Act has an accompanying code of practice on marking and labelling of AI-generated content[96] which was published in June 2026.

Further reading

- The Guidance for AI Adoption[97], developed by Gradient Institute with the National AI Centre, contains 6 essential practices for responsible AI governance and adoption, targeted towards small business owners.
- The Australian Cyber Security Centre's page on engaging with artificial intelligence[98] explains AI risks through a cybersecurity lens.
- The Australian eSafety Commissioner[99] has many useful resources for understanding online harms including those caused by AI.
- The Office of the Australian Information Commission (OAIC) provides tips on how to protect your privacy[100].
- The University of Technology Sydney's report[101] and the Australian Framework for Generative Artificial Intelligence (AI) in Schools[102] are useful resources for K-12 educators.
- For public servants, the Digital Transformation Agency has a page on using public generative AI tools safely and responsibly[103]. The NSW Government has guidance on the use of generative AI[104] for NSW Government staff.
- The Commonwealth Bank of Australia has an approachable 1-pager on staying safe with AI[105] for Australian consumers.
- The International AI Safety Report[106] (Feb 2026) contains a summary of the state of scientific research on the capabilities and risks of generative AI systems. It is authored by over 100 AI experts worldwide.
- The AI Incident Database[107] catalogues real-world harms caused by AI systems, from hallucinations in legal filings to deepfake scams and privacy breaches. It is a useful reference if you want to explore specific examples of the risks discussed in this guide.

Authors

Yaya Lu and Dr Alistair Reid (with input from the Gradient Institute team).

Acknowledgements

Gradient Institute would like to thank our expert and community reviewers for their valuable input into this guide.

This resource was created by Gradient Institute as part of a program providing science-based AI education to help Australians understand the technology, the science behind it and its impacts. This program is made possible by a grant from Google.org, Google's philanthropic arm. Google has not been involved in any way in the scoping, preparation or other process leading to this resource apart from providing overall program funding.

We'd love your feedback.

We would love to get your thoughts and feedback on this guide. If you have additional tips or resources to help people use AI safely, please send them to us. We realise that AI capabilities and risks change continually and we expect to periodically update this guide.

Write to us at info@gradientinstitute.org.

Gradient Institute is an independent not-for-profit research organisation based in Australia. We provide science-based clarity on AI and its impacts, empowering governments, industry, civil society, and the public to make informed decisions that advance a safe and responsible AI future. Find us at: gradientinstitute.org.

References

1. Employment and Training Administration. (2026). *The U.S. Department of Labor's Artificial Intelligence Literacy Framework*. U.S. Department of Labor. [online] Available at: https://www.dol.gov/sites/dolgov/files/ETA/advisories/TEN/2025/TEN%2007-25/TEN%2007-25%20(complete%20document).pdf [Accessed 12 May 2026].

2. Benton Institute for Broadband & Society. (2026). *A Guide for AI Literacy Efforts*. Benton Institute. [online] Available at: https://www.benton.org/blog/guide-ai-literacy-efforts [Accessed 12 May 2026].

3. Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S.R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S.R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M. and Perez, E. (2023). Towards Understanding Sycophancy in Language Models. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2310.13548.

4. Epoch AI. (2021). Epoch AI. [online] Available at: https://epoch.ai/.

5. Microsoft Support (2026). *Getting started with Microsoft Copilot.* [online] Microsoft. Available at: https://support.microsoft.com/en-us/microsoft-copilot/getting-started-with-microsoft-copilot [Accessed 2 Jun.. 2026].

6. Microsoft Support (2026). *Microsoft Copilot privacy controls*. [online] Available at: https://support.microsoft.com/en-us/microsoft-copilot/microsoft-copilot-privacy-controls [Accessed 29 May. 2026].

7. Joshua, C. (2026). *How to opt out of Meta AI and protect your data.* [online] Norton. Available at: https://au.norton.com/blog/ai/how-to-opt-out-of-meta-ai [Accessed 2 Jun. 2026].

8. Griffin, L. (2026). *Old habits die hard: Microsoft tries to limit our options, this time with AI.* [online] The Mozilla Blog. Available at: https://blog.mozilla.org/en/mozilla/ai/microsoft-copilot-ai-user-choice/ [Accessed 16 May 2026].

9. Electronic Privacy Information Center. (2026). *Good Luck Opting Out: Manipulative Design Patterns in Opt-Out Processes.* [online] Available at: https://epic.org/good-luck-opting-out-manipulative-design-patterns-in-opt-out-processes-2/ [Accessed 15 May. 2026].

10. Adobe Acrobat. (2025). *AI PDF Summariser: Summarise PDFs online.* [online] Adobe. Available at: https://www.adobe.com/au/acrobat/online/ai-summary-generator.html [Accessed 15 May 2026].

11. Adobe. (2024). *Known issues with generative AI features*. [online] Available at: https://helpx.adobe.com/au/acrobat/using/known-issues-gen-ai.html [Accessed 16 May 2026].

12. Milmo, D. (2025). *Man develops rare condition after ChatGPT query over stopping eating salt.* [online] The Guardian. Available at: https://www.theguardian.com/technology/2025/aug/12/us-man-bromism-salt-diet-chatgpt-openai-health-information [Accessed 2 Jun. 2026].

13. Common Sense Media. (2025). *Meta AI Risk Assessment.* [online] Available at: https://www.commonsensemedia.org/ai-ratings/meta-ai-risk-assessment [Accessed 28 May 2026].

14. McLennan, A. (2025). *AI chatbots accused of encouraging teen suicide as experts sound alarm.* [online] ABC News. Available at:

https://www.abc.net.au/news/2025-08-12/how-young-australians-being-impacted-by-ai/105630108. [Accessed 28 May 2026].

15. Charlotin, D. (2025). *AI Hallucination Cases Database.* [online] Damiencharlotin.com. Available at: https://www.damiencharlotin.com/hallucinations/. [Accessed 28 May 2026].

16. Tufan Neupane (2025). *As more lawyers fall for AI hallucinations, ChatGPT says: Check my work.* [online] Cronkite News. Available at: https://cronkitenews.azpbs.org/2025/10/28/lawyers-ai-hallucinations-chatgpt/. [Accessed 28 May 2026].

17. Tiller, N.B., Marcon, A.R., Zenone, M., Kidd, K.E., Jeukendrup, A.E., Master, Z. and Caulfield, T. (2026). *Generative artificial intelligence-driven chatbots and medical misinformation: an accuracy, referencing and readability audit.* BMJ Open, [online] 16(4), p.e112695. doi:https://doi.org/10.1136/bmjopen-2025-112695.

18. Beusekom, M. V. (2026). *AI chatbots provide poor answers to medical questions half the time, study finds.* CIDRAP. University of Minnesota. [online] Available at: https://www.cidrap.umn.edu/misc-emerging-topics/ai-chatbots-provide-poor-answers-medical-questions-half-time-study-finds. [Accessed 12 May 2026].

19. Magesh, V., Surani, F., Dahl, M., Suzgun, M., Manning, C. and Ho, D. (2024). *AI on Trial: Legal Models Hallucinate in 1 out of 6 (or More) Benchmarking Queries.* [online] Human-Centered Artificial Intelligence. Stanford University. Available at: https://hai.stanford.edu/news/ai-trial-legal-models-hallucinate-1-out-6-or-more-benchmarking-queries. [Accessed 9 May 2026].

20. Riddle, K. (2026). *ChatGPT might give you bad medical advice, studies warn.* [online] NPR. Available at: https://www.npr.org/2026/03/11/nx-s1-5744035/chatgpt-might-give-you-bad-medical-advice-studies-warn. [Accessed 12 May 2026].

21. Andoh, E. (2026). *AI chatbots and digital companions are reshaping emotional connection.* [online] American Psychological Association. Available at: https://www.apa.org/monitor/2026/01-02/trends-digital-ai-relationships-emotional-connection. [Accessed 11 May 2026]

22. Criddle, C. & Thomas, D. (2026). *'Unsettling' adverts are coming to your AI chatbot.* [online] The Australian Financial Review. Available at: https://www.afr.com/companies/media-and-marketing/unsettling-adverts-are-coming-to-your-ai-chatbot-20260216-p5o2ql. [Accessed 2 Jun. 2026]

23. Dunbar, M. (2025). *Chicago Sun-Times confirms AI was used to create reading list of books that don't exist.* [online] The Guardian. Available at: https://www.theguardian.com/us-news/2025/may/20/chicago-sun-times-ai-summer-reading-list. [Accessed 1 Jun. 2026]

24. Velvet Sundown Music. (2025). *Velvet Sundown.* [online] Otway St. Mark. Available at: https://www.velvetsundownmusic.com/. [Accessed 1 Jun. 2026]

25. Bakare, L. (2025). *An AI-generated band got 1m plays on Spotify. Now music insiders say listeners should be warned.* [online] The Guardian. Available at: https://www.theguardian.com/technology/2025/jul/14/an-ai-generated-band-got-1m-plays-on-spotify-now-music-insiders-say-listeners-should-be-warned. [Accessed 1 Jun. 2026]

26. Choi, A. and Mei, K. (2025). *What are AI hallucinations? Why AIs sometimes make things up.* [online] The Conversation. Available at:

https://theconversation.com/what-are-ai-hallucinations-why-ais-sometimes-make-things-up-242896. [Accessed 12 May 2026].

27. Sharma, M., Tong, M., Korbak, T., Duvenaud, D., Askell, A., Bowman, S.R., Cheng, N., Durmus, E., Hatfield-Dodds, Z., Johnston, S.R., Kravec, S., Maxwell, T., McCandlish, S., Ndousse, K., Rausch, O., Schiefer, N., Yan, D., Zhang, M. and Perez, E. (2023). Towards Understanding Sycophancy in Language Models. [online] arXiv.org. doi:https://doi.org/10.48550/arXiv.2310.13548.

28. Hutson, M. (2026). *AI Sycophancy: Why Chatbots Agree With You.* [online] IEEE Spectrum. Available at: https://spectrum.ieee.org/ai-sycophancy. [Accessed 12 May 2026].

29. ARTICLE 19. (2025). *Algorithmic people-pleasers: Are AI chatbots telling you what you want to hear?* [online] ARTICLE 19. Available at: https://www.article19.org/resources/algorithmic-people-pleasers-are-ai-chatbots-telling-you-what-you-want-to-hear/. [Accessed: 12 May 2026]

30. Merriam-Webster (2025). *2025 Word of the Year.* [online] Merriam-Webster. Available at: https://www.merriam-webster.com/wordplay/word-of-the-year. [Accessed 19 May 2026]

31. Semancik, A. (2026). *What is 'AI slop': OHIO AI faculty experts explain.* [online] OHIO today. Available at: https://www.ohio.edu/news/2026/05/what-ai-slop-ohio-ai-faculty-experts-explain. [Accessed 20 May 2026].

32. Claude Support. (2026). *Using incognito chats.* Anthropic. Available at: https://support.claude.com/en/articles/12260368-using-incognito-chats. [Accessed 18 Jun 2026]

33. OpenAI Help. (2026). *Memory FAQ.* OpenAI. [online] Available at: https://help.openai.com/en/articles/8590148-memory-faq#how-do-i-enable-or-disable-saved-memories. [Accessed 18 Jun 2026]

34. Office of the Australian Information Commissioner. (n.d.). *Your privacy rights.* [online] Australian Government. Available at: https://www.oaic.gov.au/privacy/your-privacy-rights. [Accessed 25 May 2026].

35. Ray, S. (2023). *Samsung Bans ChatGPT Among Employees After Sensitive Code Leak.* [online] Forbes. Available at: https://www.forbes.com/sites/siladityaray/2023/05/02/samsung-bans-chatgpt-and-other-chatbots-for-employees-after-sensitive-code-leak/. [Accessed 25 Jun. 2026]

36. McMahon, L. (2025). *Hundreds of thousands of Grok chats exposed in Google results*. [online] BBC News. Available at: https://www.bbc.com/news/articles/cdrkmk00jy0o. [Accessed 1 May 2026]

37. OpenAI. (2026). *Privacy policy*. [online] OpenAI. Available at: https://openai.com/en-GB/policies/row-privacy-policy/. [Accessed 20 May 2026]

38. Anthropic. (2026). *Anthropic Privacy Center.* [online] Anthropic. Available at: https://privacy.claude.com/en/. [Accessed 20 May 2026]

39. Microsoft Learn. (2025). *Data, Privacy, and Security for Microsoft 365 Copilot.* [online] Microsoft. Available at: https://learn.microsoft.com/en-us/microsoft-365/copilot/microsoft-365-copilot-privacy. [Accessed 20 May 2026]

40. Gemini Apps Help. (2026). *Gemini Apps Privacy Hub.* [online] Google Support. Available at: https://support.google.com/gemini/answer/13594961?hl=en. [Accessed 20 May 2026]

41. DeepSeek. (2026). *DeepSeek Privacy Policy.* [online] DeepSeek. Available at: https://cdn.deepseek.com/policies/en-US/deepseek-privacy-policy.html. [Accessed 20 May 2026]

42. Itoi, N. G. (2025). *Be Careful What You Tell Your AI Chatbot.* [online] Human-Centered Artificial Intelligence. Stanford University. Available at: https://hai.stanford.edu/news/be-careful-what-you-tell-your-ai-chatbot. [Accessed 21 May 2026].

43. Open AI Support. (2026). *How your data is used to improve model performance.* [online] Open AI. Available at: https://help.openai.com/en/articles/5722486-how-your-data-is-used-to-improve-model-performance. [Accessed 2 Jun. 2026].

44. Miller, K. (2024). *Privacy in an AI Era: How Do We Protect Our Personal Information?* [online] Human-Centered Artificial Intelligence. Stanford University. Available at: https://hai.stanford.edu/news/privacy-ai-era-how-do-we-protect-our-personal-information. [Accessed 21 May 2026].

45. Krantz, T., Jonker, A. and McGrath, A.(n.d.) *What is shadow AI?* [online] IBM Think. Available at: https://www.ibm.com/think/topics/shadow-ai. [Accessed 22 May 2026]

46. Office of the Australian Information Commissioner. (n.d.) *Tips to protect your privacy*. [online] Australian Government. Available at: https://www.oaic.gov.au/privacy/your-privacy-rights/ways-to-protect-your-privacy/tips-to-protect-your-privacy. [Accessed 20 May 2026].

47. Office of the Australian Information Commissioner. (n.d.) *Guidance on privacy and the use of commercially available AI products.* [online] Australian Government. Available at: https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-the-use-of-commercially-available-ai-products. [Accessed 20 May 2026].

48. Cecco, L. (2024). *Air Canada ordered to pay customer who was misled by airline's chatbot.* [online] The Guardian. Available at: https://www.theguardian.com/world/2024/feb/16/air-canada-chatbot-lawsuit. [Accessed 13 May 2026]

49. Charlotin, D. (2026). *AI Hallucination Cases.* [online] Damien Charlotin. Available at: https://www.damiencharlotin.com/hallucinations/. [Accessed 1 Jun 2026].

50. Melendez, S. (2026). *AI keeps inventing fake cases. Lawyers keep citing them.* [online] Scientific American. Available at: https://www.scientificamerican.com/article/why-lawyers-keep-citing-fake-cases-invented-by-ai/. [Accessed 20 May 2026].

51. Gosier, C. (2025). *Fatphobia Is Fueled By AI-Created Images, Study Finds.* [online] Fordham University Now. Available at: https://now.fordham.edu/science-and-technology/fatphobia-is-fueled-by-ai-created-images-study-finds/. [Accessed 2 Jun. 2026].

52. Leaver, T. and Srdarov, S. (2025). *'Australiana' images made by AI are racist and full of tired cliches, new study shows*. [online] The Conversation. Available at: https://theconversation.com/australiana-images-made-by-ai-are-racist-and-full-of-tired-cliches-new-study-shows-263117. [Accessed 2 Jun. 2026]

53. Grynhaum, M. M. and Mac, R. (2023). *The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work.* [online]. The New York Times. Available at: https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html. [Accessed 1 Jun. 2026].

54. Schor, Z. (2025). *Anderson v. Stability AI: The Landmark Case Unpacking the Copyright Risks of AI Image Generators.* [online] NYU Law Journal of Intellectual Property & Entertainment Law. Available at:

https://jipel.law.nyu.edu/andersen-v-stability-ai-the-landmark-case-unpacking-the-copyright-risks-of-ai-image-generators/. [Accessed 1 Jun. 2026]

55. Perrigo, B. (2023). *Exclusive: OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic.* [online] TIME. Available at: https://time.com/6247678/openai-chatgpt-kenya-workers/. [Accessed 5 Jun. 2026]

56. Finlay, L., Hooton, P. and Wallace, C. (2023). *Big tech is Ignoring the Human Cost Behind the Rise of ChatGPT.* [online]. Australian Human Rights Commission. Available at: https://humanrights.gov.au/about-us/media-centre/opinion-pieces/opinion-pieces/big-tech-ignoring-human-cost-behind-rise-chatgpt. [Accessed 1 Jun. 2026]

57. Fleury, M. and Jimenez, N. (2025). *'I can't drink the water' - life next to a US data centre.* [online] BBC News. Available at: https://www.bbc.com/news/articles/cy8gy7lv448o. [Accessed 1 Jun 2026]

58. Amherst College Library. (2026). *Generative AI: Ethics and Costs.* [online] Amherst College Library. Available at: https://libguides.amherst.edu/c.php?g=1350530&p=9969379. [Accessed 1 Jun 2026].

59. United Nations Environment Programme. (2025). *AI has an environmental problem. Here's what the world can do about it.* [online] United Nations. Available at: https://www.unep.org/news-and-stories/story/ai-has-environmental-problem-heres-what-world-can-do-about. [Accessed 1 Jun. 2026]

60. Masley, A. (n.d.). *What your chatbot use actually costs.* [online]. Andy Masley. Available at: https://www.andymasley.com/visuals/ai-prompt-footprint/. [Accessed 18 Jun 2026]

61. Vardon, M. (2026). *How much water and power will AI data centres use in Australia? Ironically, we don't have the data to know.* [online]. The Conversation. Available at: https://theconversation.com/how-much-water-and-power-will-ai-data-centres-use-in-australia-ironically-we-dont-have-the-data-to-know-284069. [Accessed 18 Jun 2026]

62. Goodier, M. (2025). *Revealed: Thousands of UK university students caught cheating using AI.* (online) The Guardian. Available at: https://www.theguardian.com/education/2025/jun/15/thousands-of-uk-university-students-caught-cheating-using-ai-artificial-intelligence-survey. [Accessed 20 May 2026]

63. Bastani, H., Bastani, O., Alp Sungu, Ge, H., Özge Kabakcı and Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences*, 122(26). doi:https://doi.org/10.1073/pnas.2422633122.

64. Slade, J. J. (2025). *How to help your students use AI without losing the learning.* [online] American Psychological Association. Available at: https://www.apa.org/ed/precollege/psychology-teacher-network/introductory-psychology/learning-artificial-intelligence. [Accessed 2 Jun. 2026]

65. Gerlich, M. (2025). AI Tools in Society: Impacts on Cognitive Offloading and the Future of Critical Thinking. Societies, [online] 15(1), p.6. doi:https://doi.org/10.3390/soc15010006.

66. Kosmyna, N. and Hauptmann, E. (2025). *Your Brain on ChatGPT: Accumulation of Cognitive Debt when Using an AI Assistant for Essay Writing Task.* [online] MIT. Available at: https://www.brainonllm.com/. [Accessed 4 May 2026]

67. Anthropic. (2026). *Anthropic Education Report: The AI Fluency Index.* [online] Anthropic. Available at: https://www.anthropic.com/research/AI-fluency-index. [Accessed 20 May 2026]

68. Lodge, J. M. and Loble, L. (2026). Artificial intelligence, cognitive offloading and implications for education. *University of Technology Sydney.* Report. https://doi.org/10.71741/4pyxmbnjaq.31302475.v2

69. Department of Education. (2025). *Australian Framework for Generative Artificial Intelligence (AI) in Schools*. [online] Australian Government. Available at: https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools. [Accessed 1 Jun 2026].

70. Mowshowitz, Z. (2025). *GPT-4o Is An Absurd Sycophant.* [online] Don't Worry About The Vase. Substack. Available: https://thezvi.substack.com/p/gpt-4o-is-an-absurd-sycophant. [Accessed 12 May 2026]

71. Hill, K. and Freedman, D. (2025). *Chatbots Can Go Into a Delusional Spiral. Here's How It Happens.* [online] The New York Times. Available at: https://www.nytimes.com/2025/08/08/technology/ai-chatbots-delusions-chatgpt.html. [Accessed 12 May 2026].

72. OpenAI. (2025). *Sycophancy in GPT-4o: what happened and what we're doing about it.* [online] OpenAI. Available at: https://openai.com/index/sycophancy-in-gpt-4o/. [Accessed 13 May 2026]

73. Zewe, A. (2026). *Personalization features can make LLMs more agreeable.* [online] MIT News. Available at: https://news.mit.edu/2026/personalization-features-can-make-llms-more-agreeable-0218. [Accessed 13 May 2026]

74. Morrin, H., Nicholls, L., Levin, M., Yiend, J., Iyengar, U., DelGuidice, F., Bhattacharya, S., Tognin, S., MacCabe, J., Twumasi, R., Alderson-Day, B. and Pollak, T.A. (2026). Artificial intelligence-associated delusions and large language models: risks, mechanisms of delusion co-creation, and safeguarding strategies. *The Lancet Psychiatry.* [online] doi:https://doi.org/10.1016/S2215-0366(25)00396-7.

75. Green, H. H. (2026). *New study raises concerns about AI chatbots fueling delusional thinking.* [online] The Guardian. Available at: https://www.theguardian.com/technology/2026/mar/14/ai-chatbots-psychosis. [Accessed 14 May 2026]

76. Boles, Sy. (2026). *What to make of 'AI psychosis'?* [online] The Harvard Gazette. Available at: https://news.harvard.edu/gazette/story/2026/04/what-to-make-of-ai-psychosis/. [Accessed 14 May 2026]

77. Australian Federal Police. (2026). *Scam alert - I'm an AFP police officer working under Operation Firestorm.* [online] Australian Federal Police. Available at: https://www.afp.gov.au/news-centre/feature/scam-alert-im-afp-police-officer-working-under-operation-firestorm. [Accessed 21 May 2026]

78. Australian Taxation Office. (2026). *Scam alerts.* [online] Australian Government. Available at: https://www.ato.gov.au/online-services/scams-cyber-safety-and-identity-protection/scam-alerts. [Accessed 21 May 2026]

79. 10 News (2025). *Victorian Couple Lose $500k Life Savings To Deepfake Eddie McGuire Scam* [online] 10 News First. YouTube. Available at: https://www.youtube.com/shorts/UaUEK28Huco [Accessed 1 Jun. 2026].

80. Brown, M. (2025). *Football and media personality Eddie McGuire used in deepfake financial advertisement.* [online] ABC News. Available at: https://www.abc.net.au/news/2025-03-31/vic-eddie-mcguire-deepfake-video-financial-scam/105115764. [Accessed 2 Jun. 2026]

81. Australian Competition and Consumer Commission. (2024). *It's a scam! Celebrities are not getting rich from online investment trading platforms.* [online] Australian Government. Available at: https://www.accc.gov.au/media-release/its-a-scam-celebrities-are-not-getting-rich-from-online-investment-trading-platforms. [Accessed 4 Jun. 2026]

82. Elliott, D. (2025). *'This happens more frequently than people realize': Arup chief on the lessons learned from a $25m deepfake crime*. [online] World Economic Forum. Available at: https://www.weforum.org/stories/2025/02/deepfake-ai-cybercrime-arup/ [Accessed 4 Jun 2026]

83. TRUȚĂ, F. (2026). *Florida Woman Loses $15K to AI Voice Scam Mimicking Daughter in Distress.* [online] Bitdefender. Available at: https://www.bitdefender.com/en-au/blog/hotforsecurity/florida-woman-loses-15k-to-ai-voice-scam-mimicking-daughter-in-distress. [Accessed 4 Jun 2026]

84. Federal Bureau of Investigation. (2024). *Public Service Announcement: Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud.* [online] Alert Number: I-120324-PSA. U.S. Department of Justice. Available at: https://www.ic3.gov/PSA/2024/PSA241203. [Accessed 3 Jun 2026]

85. Scamwatch. (2026). National Anti-Scam Centre. Australian Competition and Consumer Commission. Australian Government. [online]. Available at: https://www.scamwatch.gov.au/. [Accessed 3 Jun 2026]

86. Scamwatch. (2026). *Targeting scams: report of the National Anti-Scam Centre on scams data and activity 2025.* National Anti-Scam Centre. Australian Competition and Consumer Commission. Australian Government. [online] Available at: https://www.nasc.gov.au/reports-and-publications/targeting-scams/targeting-scams-report-of-the-national-anti-scam-centre-on-scams-data-and-activity-2025. [Accessed 3 Jun 2026]

87. Australian Cyber Security Centre. (2026). *Cyber.gov.au*. Australian Signals Directorate. Australian Government. [online]. Available at: https://www.cyber.gov.au/. [Accessed 20 May 2026]

88. Commbank Newsroom. (2026). *How good are Australians at spotting an AI-powered deepfake scam?* Commonwealth Bank of Australia. [online] Available at: https://www.commbank.com.au/articles/newsroom/2026/01/can-australians-spot-deepfake-scams.html. [Accessed 20 May 2026]

89. eSafety Commissioner. (2022). *Deepfake trends and challenges - position statement.* eSafety Commissioner. Australian Government. [online]. Available at: https://www.esafety.gov.au/sites/default/files/2022-01/Deepfake-position-statement%20_v2.pdf. [Accessed 20 May 2026]

90. Thomson, T.J. (2025). *Can you tell the difference between real and fake news photos?.* RMIT University. The Conversation. [online]. Available at: https://theconversation.com/can-you-tell-the-difference-between-real-and-fake-news-photos-take-the-quiz-to-find-out-253539. [Accessed 3 Jun 2026]

91. Workman, M., Martino, M. and Carter, L. (2026). *Foreign Facebook accounts using AI Pauline Hanson to manipulate Australians.* ABC NEWS Verify. [online]. Available at: https://www.abc.net.au/news/2026-03-11/foreign-fake-news-pauline-hanson-one-nation/106436702. [Accessed 1 Jul 2026]

92. Wrona, A. (2025). *9 claims we've fact-checked about Peyton Manning.* Snopes. [online]. Available at: https://www.snopes.com/collections/peyton-manning-claims-collection/. [Accessed 4 Jun 2026]

93. Liles, J. (2025). *Texas floods: Myriad of misleading claims besiege Tom Brady in aftermath.* Snopes. [online]. Available at: https://www.snopes.com/fact-check/texas-floods-tom-brady/. [Accessed 4 Jun 2026]

94. eSafety Commissioner. (2026). *Deepfake trends and challenges - position statement.* eSafety Commissioner. Australian Government. [online] Available at: https://www.esafety.gov.au/industry/tech-trends-and-challenges/deepfakes. [Accessed 3 Jun 2026].

95. Miao, F., UNESCO, Shiohira, K. and Lao, N. (2026). *AI competency framework for students.* UNESCO. [online]. Available at: https://www.unesco.org/en/articles/ai-competency-framework-students. [Accessed 21 May 2026].

96. European Commission. (2026). *Code of Practice on Transparency of AI-Generated Content.* European Commission. [online]. Available at: https://digital-strategy.ec.europa.eu/en/policies/code-practice-ai-generated-content. [Accessed 21 May 2026]

97. National AI Centre. (2026). *Essential AI practices.* National AI Centre. Australian Government. [online]. Available at: https://www.ai.gov.au/staying-safe-and-responsible/essential-ai-practices. [Accessed 3 Jun 2026]

98. Australian Cyber Security Centre. (2024). *Engaging with artificial intelligence.* Australian Signal Directorate. Australian Government. Available at: https://www.cyber.gov.au/business-government/secure-design/artificial-intelligence/engaging-with-artificial-intelligence. [Accessed 4 May 2026]

99. eSafety Commissioner. (2026). *eSafety Homepage.* eSafety Commissioner. Australian Government. [online]. Available at: https://www.esafety.gov.au/sites/default/files/2022-01/Deepfake-position-statement%20_v2.pdf. [Accessed 1 Jul 2026]

100. Office of the Australian Information Commissioner. (2026). *Tips to protect your privacy.* Office of the Australian Information Commissioner. Australian Government. Available at: https://www.oaic.gov.au/privacy/your-privacy-rights/ways-to-protect-your-privacy/tips-to-protect-your-privacy. [Accessed 4 Jun. 2026]

101. Lodge, J.M., Loble, L. (2026). *Artificial intelligence, cognitive offloading and implications for education.* University of Technology Sydney. Paul Ramsay Foundation grant. [online] Available at: https://figshare.uts.edu.au/articles/report/Artificial_intelligence_cognitive_offloading_and_implications_for_education/31302475?file=62363005. [Accessed 21 May 2026]

102. Department of Education. (2025). *Australian Framework for Generative Artificial Intelligence (AI) in Schools.* Department of Education. Australian Government. [online] Available at: https://www.education.gov.au/schooling/resources/australian-framework-generative-artificial-intelligence-ai-schools. [Accessed 20 May 2026]

103. Digital.gov.au. (2026). *Staff guidance on public generative AI.* Digital.gov.au. Australian Government. [online]. Available at: https://www.digital.gov.au/policy/ai/staff-guidance-public-generative-ai. [Accessed 12 May 2026]

104. Digital NSW. (2026). *Generative AI: basic guidance.* NSW Government. [online]. Available at: https://www.digital.nsw.gov.au/policy/artificial-intelligence/generative-ai-basic-guidance. [Accessed 18 Jun 2026].

105. Commonwealth Bank of Australia. (2025). *Tips to Stay Safe with AI.* Commonwealth Bank of Australia. [online] Available at: https://www.commbank.com.au/content/dam/commbank-assets/about-us/2025/Staying-safe-with-AI.pdf. [Accessed 13 May 2026].

106. Bengio, Y., Clare, S., Prunkl, C., Murray, M., Andriushchenko, M., Bucknall, B., Bommasani, R., Casper, S., Davidson, T., Douglas, R., Duvenaud, D., Fox, P., Gohar, U., Hadshar, R., Ho, A., Hu, T., Jones, C., Kapoor, S., Kasirzadeh, A., Manning, S., Maslej, N., Mavroudis, V., McGlynn, C., Moulange, R., Newman, J., Ng, K.Y., Paskov, P., Rismani, S., Sastry, G., Seger, E., Singer, S., Stix, C., Velasco, L., Wheeler, N., Acemoglu, D., Conitzer, V., Dietterich, T.G., Felten, E.W., Heintz, F., Hinton, G., Jennings, N., Leavy, S., Ludermir, T., Marda, V., Margetts, H., McDermid, J., Munga, J., Narayanan, A., Nelson, A., Neppel, C., Ramchurn, S.D., Russell, S., Schaake, M., Schölkopf, B., Soto, A., Tiedrich, L., Varoquaux, G., Yao, A., Zhang, Y.-Q., Aguirre, L.A., Ajala, O., Albalawi, F., AlMalek, N., Busch, C., Collas, J., de Carvalho, A.C.P.L.F., Gill, A., Hatip, A.H., Heikkilä, J., Johnson, C., Jolly, G., Katzir, Z., Kerema, M.N., Kitano, H., Krüger, A., Lee, K.M., López Portillo, J.R., McLysaght, A., Molchanovskyi, O., Monti, A., Nemer, M., Oliver, N.,

Pezoa, R., Plonk, A., Ravindran, B., Riza, H., Rugege, C., Sheikh, H., Wong, D., Zeng, Y., Zhu, L., Privitera, D. and Mindermann, S. (2026) *International AI Safety Report 2026*. DSIT 2026/001. [online]. Available at: https://internationalaisafetyreport.org [Accessed: 2 June 2026].

107. Partnership on AI. (2026). *Welcome to the AI Incident Database.* Partnership on AI. [online]. Available at: https://incidentdatabase.ai/. [Accessed 16 May 2026].