

Australian Framework for Generative Artificial Intelligence in Schools

Consultation paper

Contents

| | |
|--|----|
| Australian Framework for Generative Artificial Intelligence in Schools | 1 |
| Background to consultation | 3 |
| Draft Australian Framework for Generative Artificial Intelligence in Schools | 4 |
| Context and overview | 4 |
| Australian Framework for Generative Artificial Intelligence in Schools | 6 |
| Consultation questions | 8 |
| 1. Teaching and learning | 8 |
| 2. Human and social wellbeing | 9 |
| 3. Transparency | 10 |
| 4. Fairness | 11 |
| 5. Accountability | 13 |
| 6. Privacy and security | 14 |
| How to provide your feedback | 16 |

Background to consultation

This paper has been developed by the Education Ministers Artificial Intelligence in Schools Taskforce (AI Taskforce) to support consultation on the draft Australian Framework for Generative Artificial Intelligence in Schools (the draft Framework).

It includes the draft Framework, which sets out the core elements and principles that will guide education systems, schools and teachers in using generative AI safely and ethically, to improve teaching and learning, lift student outcomes, and reduce the administrative and workload burden in schools.

To support understanding and discussion of the intention behind the core elements and principles of the draft Framework, the consultation paper provides additional explanation of each principle, and key questions for all education stakeholders.

Your feedback on the draft Framework will help refine the final Australian Framework for Generative Artificial Intelligence in Schools for consideration by Education Ministers before the end of 2023.

Draft Australian Framework for Generative Artificial Intelligence in Schools

Context and overview

The Australian Framework for Generative Artificial Intelligence in Schools (the Framework) seeks to guide the responsible and ethical use of generative AI tools in ways that benefit students, schools, and society.

Purpose

Generative Artificial Intelligence (AI) technology has transformative potential to improve teaching and learning in Australian schools. For students, it can help to synthesise complex ideas, create personalised content, and provide targeted and instantaneous feedback. For teachers, it can tailor instruction, simplify planning, and streamline administrative tasks. The growing accessibility and sophistication of generative AI tools provides opportunities to develop human-like generated text and rich multimedia content in a way that has not previously been possible.

To fully harness the potential of high quality and safe generative AI, schools will need to understand and appropriately manage a range of interconnected privacy, security, and ethical risks. These include the potential for errors and algorithmic bias in generative AI content; the misuse of personal or confidential information; and the use of generative AI for inappropriate purposes, such as to discriminate against individuals or groups, or to undermine the integrity of student assessments.

In February 2023, Education Ministers agreed that responding to the risks and harnessing opportunities for Australian schools and students arising from generative AI technologies is a national education priority. Ministers agreed to develop an evidence-informed, best practice framework for Australian schools.

About the Framework

The Framework is an evolving document that articulates the principles underpinning the safe and ethical use of generative AI tools in Australian schools. It has been designed as a principles-based framework that provides the necessary foundation for the development of practical guidance to schools on the safe and ethical use of generative AI.

The Framework encompasses 6 core elements of teaching and learning, human and social wellbeing, transparency, fairness, accountability, and privacy and security, with 22 underlying principles that are relevant to the responsible and ethical practice of teaching, learning, and working in or with Australian schools. The Framework complements [Australia's Artificial Intelligence Ethics Framework](#), establishing principles that address the unique challenges associated with using generative AI tools in educational environments.

The Framework aligns with the two goals of the [Alice Springs \(Mparntwe\) Education Declaration](#): promoting equity and excellence in education, and enabling all young Australians to become confident and creative individuals, successful lifelong learners, and active and informed members of the community. The Framework supports the realisation of these goals, fostering innovation in teaching and learning while upholding the principles of inclusivity, fairness, and human rights.

Target audience

The Framework is being developed for Australian education systems and schools, with each principle designed to guide the practice of schools and their staff. Additionally, the Framework will help guide teachers in instructing students how to use generative AI tools, with the principles of the Framework setting clear expectations for how students can utilise these tools in safe, responsible and ethical ways. State and territory policymakers and AI developers may also use this framework to understand what best-practice implementation of generative AI tools in a school context can look like.

Development of the Framework

Education Ministers established a national Taskforce to develop this draft Framework for consultation.

The fast-moving pace of technological development in generative AI necessitates a regular window for review. This Framework will be reviewed within 12 months of publication and every 12 months thereafter, or as needed.

What is generative AI?

Generative AI is a type of computer-based model that can generate new content, such as text, images, audio, and video. It utilises machine learning, a process in which it is trained to recognise complex patterns in large data sets, producing outputs that can closely resemble human-generated content, without further explicit programming.

Australian Framework for Generative Artificial Intelligence in Schools

| Core elements | Principles |
|--|---|
| <p>1. Teaching and learning <i>Generative AI tools are used to enhance teaching and learning.</i></p> | <p>1.1 Impact: the use of generative AI tools has a positive impact on teaching and learning outcomes.</p> <p>1.2 Instruction: schools engage students in learning about generative AI tools and how they work, including their potentials limitations and biases, and deepening this learning as student usage increases.</p> <p>1.3 Human cognition: generative AI tools are leveraged and harnessed to support and not restrict human thought and experience.</p> |
| <p>2. Human and social wellbeing <i>Generative AI tools are used to benefit all members of the school community.</i></p> | <p>2.1 Wellbeing: the use of generative AI tools has a positive impact on, and does not harm, wellbeing of all members of the school community.</p> <p>2.2 Diversity of perspectives: generative AI tools expose users to diverse ideas and perspectives, avoiding the reinforcement of existing biases.</p> <p>2.3 Human rights: the use of generative AI respects human rights and safeguards the autonomy of individuals, especially children.</p> |
| <p>3. Transparency <i>Students, teachers, and schools understand how generative AI tools work, and when and how these tools are impacting them.</i></p> | <p>3.1 Information: clear, comprehensive, and developmentally appropriate information is provided about any generative AI tool that is being used.</p> <p>3.2 Disclosure: people are informed when they are significantly impacted or being engaged by generative AI.</p> <p>3.3 Explainability: generative AI tools are explainable, so that humans can understand the reasoning behind the AI model's outputs or predictions, which includes understanding when algorithmic bias has influenced AI outputs.</p> |
| <p>4. Fairness <i>Generative AI tools are used in ways that are accessible, fair and respectful.</i></p> | <p>4.1 Accessibility: generative AI tools are used in ways that are inclusive, accessible, and equitable for people with disability and from diverse backgrounds.</p> <p>4.2 Non-discrimination: the use of generative AI tools minimises opportunities for, and counters, discrimination against individuals, communities, or groups.</p> <p>4.3 Academic integrity: when used in assessments, generative AI tools provide a fair and unbiased evaluation of students' performance, skills, and knowledge.</p> <p>4.4 Cultural and intellectual property: when using generative AI tools, schools respect the cultural rights of various cultural groups; in particular, Indigenous Cultural and Intellectual Property (ICIP) rights.</p> <p>4.5 Copyright: when using generative AI tools, schools are aware of applicable copyright rights and obligations.</p> |
| <p>5. Accountability <i>Generative AI tools are used in ways that are open to challenge and retain human agency and accountability for decisions.</i></p> | <p>5.1 Human accountability: decisions remain in human control with clear human accountability.</p> <p>5.2 Reliability: generative AI tools are well understood before they are used, and reliably operate in accordance with their intended purpose.</p> <p>5.3 Monitoring: schools regularly monitor the generative AI tools they are using and their impact on students and teachers.</p> <p>5.4 Contestability: individuals (e.g., students, parents, staff) that are significantly impacted by a generative AI tool are able to challenge the use or outputs of the tool, and any decisions informed by the tool.</p> |
| <p>6. Privacy and security <i>Students and others using generative AI tools have their privacy and data protected.</i></p> | <p>6.1 Privacy and data protection: generative AI tools are used to respect and uphold privacy and data rights and comply with Australian law.</p> <p>6.2 Privacy disclosure: students, parents, and stakeholders are proactively informed about how data will be collected, used, and shared while using a generative AI tool.</p> |

| | |
|--|---|
| | <p>6.3. Protection of student inputs: generative AI tools are used in ways that protect inputs by or about students, such as typed prompts, uploaded multimedia, or other data, via best practice privacy-preserving data sharing methods.</p> <p>6.4. Cybersecurity and resilience: robust cyber-security measures are implemented to protect the integrity and availability of school infrastructure, generative AI tools, and associated data.</p> |
|--|---|

Consultation on the draft Framework

You are invited to provide feedback on the 6 core elements and the principles outlined in the draft Framework.

This section of the consultation paper supplies more information about each of the draft principles. Your response to the key consultation questions will help refine the final version of the Framework.

Key consultation questions

- **Will these core elements and principles help to guide Australian school systems, teachers and school administrators in using generative AI safely and in ways that support better education outcomes?**
- **What changes to the core elements and principles are needed to provide a clear framework that will support the safe and effective use of generative AI in schools?**

Understanding the core elements and principles

1. Teaching and learning

Generative AI tools are used to enhance teaching and learning

1.1 Impact

The use of generative AI tools has a positive impact on teaching and learning outcomes.

This principle seeks to ensure that generative AI tools are used to improve teaching and learning outcomes, and that these outcomes clearly outweigh risks or negative implications.

It is important for schools and education systems to learn from the emerging evidence-base about where, when, how and why the use of generative AI is most effective.

1.2 Instruction

Schools engage students in learning about generative AI tools and how they work, including their potential limitations and biases, and deepening this learning as student usage increases.

This principle aims to ensure that students learn about generative AI systems so that they can use them effectively, avoid harms and keep up to date with the latest technological developments.

Learning about generative AI systems will enable students to use them effectively, avoid harm, and keep up with the latest technology. Understanding AI supports students to develop awareness of

their individual responsibility and accountability, and to improve their ability to use the technology in responsible and effective ways.

1.3 Human cognition

Generative AI tools are leveraged and harnessed to support and enhance rather than restrict human thought and critical thinking.

This principle aims to ensure that generative AI is used to enhance human skills such as critical thinking, judgement, and reasoning, and that these skills are not undermined or restricted through the use of or reliance on generative AI.

The use of generative AI in the classroom will need to be balanced and primarily used to enhance, augment, or complement human skills. Schools could enhance student thought and critical thinking by teaching metacognitive strategies, such as promoting self-questioning during learning, encouraging problem-solving approaches, and teaching how to evaluate information for credibility and relevance.

2. Human and social wellbeing

Generative AI tools are used to benefit all members of the school community

2.1 Wellbeing

The use of generative AI tools has a positive impact on, and does not harm, the wellbeing of all members of the school community.

This principle aims to ensure the development and use of generative AI in schools has clear benefits to the wellbeing of users and those impacted by its use.

Schools could support students to think carefully about the intended and unintended consequences that could arise from the use of generative AI before it is used, and by understanding the impact of these tools on the wellbeing of members of the school community.

2.2 Diversity of perspectives

Generative AI tools expose users to diverse ideas and perspectives, avoiding the reinforcement of existing biases.

This principle aims to ensure that generative AI tools are used in a way that exposes students to a range of views and ideas. This is essential to building compassion and critical thinking and supporting diversity in education settings.

The outputs of generative AI can be inaccurate or biased – even when expressed in confident language. Generative AI also responds to the preferences of users, which may result in the adoption of dominant ideas and remove more diverse views, creating ‘echo chambers’ and reinforcing biases.

Where generative AI is used in schools, it is important students continue to have exposure to a diverse range of ideas and perspectives. Developers and users should also be aware of, and take reasonable steps to manage, potential bias in generative AI.

2.3 Human rights

The use of generative AI respects human rights and safeguards the autonomy of individuals, especially children.

This principle aims to preserve human autonomy and ensure the development and use of generative AI respects the rights of individuals and fosters an environment of respect and dignity. Special attention is given to children, recognising their particular vulnerabilities and the need to protect their rights in all circumstances.

The use of generative AI should complement or augment human abilities and ensure humans maintain control over the use of generative AI, while protecting student dignity, autonomy and wellbeing.

Generative AI development processes are not well understood and this presents a range of risks, including potentially unethical practices that may infringe on workers' rights or disregard intellectual property laws. Developers of generative AI should be expected to source and build their datasets, products and services ethically and be transparent about their practices.

3. Transparency

Students, teachers, and schools understand how generative AI tools work, and when and how these tools are impacting them

3.1 Information

Clear, comprehensive, and developmentally appropriate information is provided about any generative AI tool that is being used.

This principle highlights the need for all users of generative AI to be informed about the function of the generative AI tool and why it is being used in the classroom.

AI is embedded in many everyday applications, including in the tools teachers use in classrooms and in the systems used by schools. All members of the school community should understand how and why generative AI is being used, and what outcomes are intended. In turn, developers can ensure that plain-English information about the workings and expected use of their products is available to consumers.

3.2 Disclosure

People are informed when they are significantly impacted or being engaged by generative AI.

This principle aims to ensure that people have been informed when generative AI is being used and that they understand the implications of its use.

The use of generative AI is not always obvious to users. When generative AI is used in schools or any other education purpose, all users are provided with easy to understand information about the tool, including any risks, and why it is being used. Disclosure is important for accountability of school systems, so people know how their data is being used and can contest its use or outputs.

3.3 Explainability

Generative AI tools are explainable, so that humans can understand the reasoning behind the AI model's outputs or predictions, which includes understanding when algorithmic bias has influenced AI outputs.

This principle aims to ensure that generative AI tools are clearly explained, including how they operate, and how input data influences the model's outputs. This includes an explanation of how the data a generative AI tool is trained on can bias its outputs.

“Explainability”, which focuses on understanding the reasoning behind AI-generated decisions and content in human terms, should be prioritised even when interpretability, the ability to present the inner workings of AI models and identify the direct relationship between inputs and outputs, is not possible.

4. Fairness

Generative AI tools are used in ways that are accessible, fair and respectful

4.1 Accessibility

Generative AI tools are used in ways that are inclusive, accessible, and equitable for people with disability and from diverse backgrounds.

This principle aims to ensure the use of generative AI supports and responds to the diverse needs, backgrounds, experiences and knowledge of all students, especially Aboriginal and Torres Strait Islander students, those with disability and those from diverse backgrounds.

Accessibility and equity should be key considerations when choosing and using generative AI tools. If used effectively in the classroom, generative AI tools have the potential to help teachers to better support the unique learning needs of students.

4.2 Non-discrimination

The use of generative AI tools minimises opportunities for, and counters, discrimination against individuals, communities, or groups.

This principle aims to ensure that the use of generative AI tools does not result in discrimination against individuals, communities, or groups.

The Large Language Models (LLMs) used in generative AI have been trained on information that includes embedded biases and can produce outputs that are discriminatory. The use of generative AI

must comply with Australian anti-discrimination laws and protect students and other members of the school community from harm.

4.3 Academic integrity

When used in assessments, generative AI tools provide a fair and unbiased evaluation of students' performance, skills, and knowledge.

This principle aims to ensure schools consider the impact generative AI tools can have on student assessments and take reasonable steps to preserve academic integrity.

Generative AI can be used to support student learning and assessment. It can also be used by students to generate answers for exams and assignments. Consideration will need to be given to identifying and responding to inappropriate use of AI to generate content. Assessments may need to be modified to avoid or use generative AI tools, so that they continue to ensure outputs will be a fair and unbiased evaluation of students' performance, skills, and knowledge.

4.4 Cultural and intellectual property

When using generative AI tools, schools respect the cultural rights of various cultural groups; in particular, Indigenous Cultural and Intellectual Property (ICIP) rights.

This principle aims to ensure generative AI is used in a way that protects the cultural and intellectual property rights of various groups.

Cultural and intellectual property rights protect cultural heritage including ideas, materials, and traditional practices. While Australian intellectual property laws only protect some forms of cultural and intellectual property rights, the acknowledgement of these rights supports self-determination and cultural integrity and demonstrates respect.

Schools can support increased awareness of the potential to infringe, inadvertently or otherwise, culture and intellectual property rights when using generative AI, and the importance of the respectful and lawful use of visual imagery, spoken or written ideas, and physical materials.

4.5 Copyright

When using generative AI tools, schools are aware of applicable copyright rights and obligations.

This principle aims to ensure that schools are aware of copyright issues in relation to the use of generative AI tools.

The use of generative AI in schools is still new, and it is possible that in some circumstances its use may breach existing copyright obligations. Generative AI systems are trained on content that may be protected by copyright. It is important schools have access to information about copyright issues in relation to the use of generative AI.

5. Accountability

Generative AI tools are used in ways that are open to challenge and retain human agency and accountability for decisions

5.1 Human accountability

Decisions remain in human control with clear human accountability.

This principle aims to ensure that generative AI is always used under human supervision with clear accountability for its use and outputs. Where generative AI is used by students or with student data it is essential that there are appropriate safeguards.

Decision-making processes should always involve human judgement and responsibility. Ensuring human accountability may include making users aware of their responsibilities and keeping records of the use of generative AI.

5.2 Reliability

Generative AI tools are well understood before they are used, and reliably operate in accordance with their intended purpose.

This principle aims to ensure that generative AI systems are used and operate reliably. There is little benefit in implementation of generative AI that does not do what is required or does not produce reliable outcomes.

Understanding how generative AI products will perform prior to implementation enables effective judgements to be made about their operation. This helps schools and the broader community to have confidence that generative AI used in school systems will operate in the way they expect and will deliver real benefits.

Developers should provide information about the reliability of generative AI products, including providing evidence that products are fit for purpose.

5.3 Monitoring

Schools regularly monitor the generative AI tools they are using and their impact on students and teachers.

This principle aims to ensure that schools monitor the impact of the generative AI tools they are using. The aim is to ensure the tools are working as expected, achieving the desired outcomes, not causing unforeseen negative impacts.

Regular monitoring and review of generative AI tools and their impact may also provide for continuous improvement and potentially support further innovation. Metrics that could be monitored in relation to the use of generative AI include the impact on academic achievement,

student and teacher wellbeing, teacher productivity, and even teacher retention. This data can support schools in choosing new products, updating or maintaining generative AI guidelines, and protecting students and teachers.

5.4 Contestability

Individuals (e.g., students, parents, staff) that are significantly impacted by a generative AI tool are able to challenge the use or outputs of the tool, and any decisions informed by the tool.

This principle aims to uphold best practice in the use of generative AI by allowing humans to intervene when the products or outputs, including decisions, may be inaccurate or wrong.

Generative AI can produce results or outputs that are inaccurate or wrong. Fairness, transparency and accountability in using generative AI cannot be achieved without ensuring that individuals have visibility and contestability. This includes developing appropriate processes to allow all members of the school community to contest the use and outputs of generative AI where possible.

This is important even when generative AI systems are performing as expected.

6. Privacy and security

Students and others using generative AI tools have their privacy and data protected

6.1 Privacy and data protection

Generative AI tools are used to respect and uphold privacy and data rights and comply with Australian law.

This principle aims to ensure that the use of generative AI aligns with public expectations about the importance of data privacy and the law.

Some uses of generative AI have the potential to match data inputs with individuals, breaching data privacy. There are also concerns about the safety and storage of data inputs into generative AI tools.

Ensuring understanding and compliance with Australian data privacy laws will minimise the risk of a data privacy breach, such as a student or schools data being compromised due to intentional or accidental misuse.

6.2 Privacy disclosure

Students, parents, and stakeholders are proactively informed about how data will be collected, used, and shared while using a generative AI tool.

This principle aims to support the understanding of how data is used, collected and shared when engaging with generative AI in education systems and schools. This understanding is important to building confidence about the use of generative AI in schools, including in school administration.

When generative AI is used in education systems and schools, key stakeholders such as parents and students will need to be informed about how data is collected and the data protection safeguards in place.

6.3 Protection of student inputs

Generative AI tools are used in ways that protect inputs by or about students, such as inputs, for instance via typed prompts, uploaded multimedia, or other data, via best practice privacy-preserving data sharing methods.

This principle aims to ensure that personal student information is protected via appropriate privacy process, such as restricting student inputs from being shared with others, or used to train the AI model. This is particularly important where student inputs might be used for analysis or other purposes outside of classroom use.

There may be instances where teachers need access to identified data, for example where student input or outputs are being used as part of classroom collaboration, teaching or assessment. Education systems can support schools and educators to recognise instances where it is important that student data is protected.

6.4 Cybersecurity and resilience

Robust cybersecurity measures are implemented to protect the integrity and availability of school infrastructure, generative AI tools, and associated data.

This principle aims to ensure that education systems and schools implement robust cybersecurity measures to protect against any risks that may arise from the use of generative AI, including protecting associated data.

Through robust cybersecurity measures, education systems and schools will protect against any risks that may arise from the use of generative AI, including protecting associated data.

How to provide your feedback

We invite your feedback on the draft Australian Framework for Artificial Intelligence in Schools.

Consultation questions

- **Will these core elements and principles help to guide Australian school systems, teachers and school administrators in using generative AI safely and in ways that support better education outcomes?**
- **What changes to the core elements and principles are needed to provide a clear framework that will support the safe and effective use of generative AI in schools?**

Submit your feedback

Please provide your feedback by 16 August 2023 via the following survey link:

<https://forms.office.com/r/ERDE4fTNyb>

If you wish to provide additional feedback, please send an email to our mailbox:

AISecretariat@det.nsw.edu.au