GENERAL GUIDELINES FOR THE USE OF GENERATIVE AI IN ASSESSMENT TASKS

Preamble:

The general advice for the use of Generative Artificial Intelligence (GAI) in schools is constantly changing and being rewritten as new developments in GAI are reached. As of late 2023, The Department of Education (DoE) is yet to release its final guidelines for the use of AI in Assessment Tasks.

Consult the following subheadings for general guidance:

- The Draft Australian Framework for Generative Artificial Intelligence in Schools
- The use of GAI detector tools
- The DoE's guidelines regarding use of generative AI (ChatGPT)
- The use of online document version histories and file properties to determine authenticity
- The *HSC: All My Own Work* program as guidance for the use of GAI in assessment tasks

Draft Australian Framework for Generative Artificial Intelligence in Schools

The Australian Framework for Generative Artificial Intelligence in Schools is in its consultation phase in late 2023. It should be consulted below to ensure GAI is used ethically and morally in the context of the teaching and learning requirements of all courses and their outcomes.

Draft Australian Framework for Generative Artificial Intelligence in Schools (2023)

Core Elements	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. **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. **1.3 Human cognition:** generative AI tools are leveraged and harnessed to support and not restrict human thought and experience.
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, wellbeing of all 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. **2.3 Human rights:** the use of generative AI respects human rights and safeguards the autonomy of individuals, especially children.
3. Transparency Students, teachers, and schools understand how generative AI tools work, and when and how these	**3.1 Information:** clear, comprehensive, and developmentally appropriate information is provided about any generative AI tool that is being used. **3.2 Disclosure:** people are informed when they are significantly impacted or being engaged by generative AI. **3.3 Explainability:** generative AI tools are explainable, so that humans can understand the reasoning behind the AI model's outputs or predictions,

tools are impacting them.	which includes understanding when algorithmic bias has influenced AI outputs.
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. **4.2 Non-discrimination:** the use of generative AI tools minimises opportunities for, and counters, discrimination against individuals, communities, or groups. **4.3 Academic integrity:** when used in assessments, generative AI tools provide 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. **4.5 Copyright:** when using generative AI tools, schools are aware of applicable copyright rights and obligations.
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. **5.2 Reliability:** generative AI tools are well understood before they are used, and reliably operate in accordance with their intended purpose. **5.3 Monitoring:** schools regularly monitor the generative AI tools they are using and their impact on 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.
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. **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. **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. **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.

The Use of Generative AI Detector Tools

AI detection tools determine whether a piece of writing has been generated by an AI bot by analysing various linguistic and structural patterns that are typical of AI-generated content. Teachers may access a variety of AI detection tools to assist in the process of determining the authenticity of student work within their course. Generative AI detection tools work by identifying and analysing the following components of a piece of submitted work:

1. **Pattern Recognition:** AI detection tools are trained on a diverse dataset of human-written and AI-generated text. They analyse patterns in sentence structure, vocabulary usage, syntax, coherence, and other linguistic features specific to AI-generated content.
2. **Syntax and Structure Analysis:** AI-generated content often exhibits certain syntax and structural patterns that distinguish it from human-written content. AI detection tools can identify these patterns, such as unnatural sentence flow or the overuse of specific phrases or terms.
3. **Statistical Analysis:** AI detection tools employ statistical methods to analyse the frequency and distribution of words, phrases, and sentence lengths. AI-generated content may display statistical anomalies or repetitive patterns not typically found in human writing.
4. **Contextual Understanding:** AI detection tools consider the context and coherence of the text to determine if it aligns with human understanding and logic. They evaluate whether the content follows a logical narrative and if it integrates knowledge and context appropriately.
5. **Language Model Analysis:** AI detection tools often use language models, including those based on neural networks like GPT (Generative Pre-trained Transformer), to assess the text. They compare the writing style and characteristics against known AI-generated language models.
6. **Training and Machine Learning:** AI detection tools are trained using machine learning algorithms on a labelled dataset that includes both AI-generated and human-written texts. The tool learns to differentiate between the patterns and features unique to AI-generated content and human-authored content.
7. **Anomaly Detection:** AI detection tools use anomaly detection techniques to identify deviations from expected patterns found in human writing. Unusual linguistic choices, non-standard grammar, or inconsistencies may indicate AI generation.
8. **Comparative Analysis:** AI detection tools may compare the analysed text with a reference corpus of known AI-generated text to establish a similarity score. Higher similarity scores suggest that the text may have been generated by an AI bot.
9. **Feedback Loops and Improvement:** AI detection tools continuously learn and improve through feedback loops. As they encounter new AI-generated patterns, they integrate this knowledge to enhance their ability to identify AI-generated content more accurately.

By employing a combination of pattern recognition, statistical analysis, language model evaluation, and machine learning techniques, AI detection tools can determine the likelihood that a given piece of writing has been generated by an AI bot rather than a human. Teachers may use these tools to assist in their decision-making process to determine whether a student's submission has been constructed either wholly or partially by generative AI.

The Department of Education's Guidelines Regarding the Use of Generative AI

The department continues to restrict student access to generative AI applications on devices within department networks. As part of its commitment to duty of care for all its students, the decision to restrict access is due to the lack of reliable safeguards preventing these tools from exposing students to potentially explicit and harmful content. While students have access to generative AI tools outside of the context of the classroom, the following guidelines are recommended for safe, ethical usage.

Safeguard Personal Information

When using generative AI tools for teaching and learning, it is essential to manage personal information effectively and ensure proper de-identification processes are in place.

What is Personal Information?

Personal information is any information that can be used to identify an individual directly or indirectly. It could be a student's name, address, class, school, family details, fingerprints or a combination of information from which a student, teacher or other individual can be identified. The information can be recorded in paper files, electronic records, video recordings and photographs.

De-identification of Personal Information

De-identification involves removing, anonymising, or masking personal information so that it can no longer be used to identify an individual. This process is crucial to protect privacy.

The following techniques will help you safeguard sensitive information while maintaining the utility of the data for analysis and learning purposes.

- Data anonymisation: Replace or alter personal information in the text using information that cannot be linked to the individual.
- Data masking: Replace personal information in the text, such as email addresses or phone numbers, with fictional data that retains the same format.
- Include only relevant data: Choose only information that is relevant.

Consulting Document, Version Histories and File Properties to Determine Authenticity

Google – Google Docs, Googles Slides, Google Sheets etc.

Consulting document version histories in Google (e.g., Google Docs) can help determine the authenticity of a piece of submitted work by providing insights into the editing and revision process. Teachers may access these sources of information to determine whether a piece of submitted work has been written entirely by a student themselves. Teachers may do this by consulting any of the following options below in a piece of work:

1. **Accessing Version History:**
    - "See version history" allows the teacher to view all saved versions of the document.
2. **Reviewing Revisions:**
    - Examine the list of versions to see the edits, additions, or deletions made at each stage.
    - Click on a specific version to view the document as it existed during that revision.
3. **Analysing Changes:**
    - Look for substantial changes or alterations that may significantly impact the content's meaning or authenticity.

- Pay attention to patterns of editing, multiple edits within a short timeframe, or sudden and significant modifications.

4. **Comparing Versions:**
   - Use the "Compare" feature to view differences between two selected versions.
   - This allows you to see precisely what was changed between the versions, aiding in identifying any suspicious alterations.
5. **Checking Timestamps:**
   - Review the timestamps associated with each version to understand when changes were made.
   - Quick edits and large dumps of text in a short space of time can be suspicious.
6. **Verifying Consistency:**
   - Assess the consistency of the writing style, formatting, and tone throughout different versions.
   - Inconsistent or abrupt changes may raise concerns about the authenticity of the document.
7. **Examining Collaborators:**
   - Check the collaborators listed during each version to identify individuals who contributed to the document.
   - Ensure that all contributors are authorised and expected collaborators when completing group work.

By leveraging the version history feature in Google Docs, teachers can closely examine the editing trail of a document, aiding in the assessment of its authenticity and adherence to academic integrity.

Microsoft – Microsoft Word, Microsoft PowerPoint, Microsoft Excel etc.

Checking the file properties of Word Docs, PowerPoints, and other Microsoft tools can provide valuable information that can help teachers to assess the authenticity of a piece of work, particularly by examining the total edit time and the number of revisions made to a file. Teachers may do this by consulting the file properties and then any of the following options below regarding a piece of work:

1. **Total Edit Time:**
   - Looking for the "Total editing time" displays the cumulative time spent editing the document.
   - Consider if the total edit time aligns with the expected effort for creating the piece of work.
2. **Number of Revisions:**
   - Analyse the number of revisions and compare it to what is reasonable for the scope and complexity of the work.

3. **Creation and Modification Dates:**
    - Check the creation date and the date of the last modification.
    - Ensure that the creation date aligns with when the work was assigned or initiated.
4. **Author Information:**
    - If available, review the author or contributor information in the file properties.
    - Verify that the listed authors and contributors are consistent with those expected to have worked on the document if the task was for a group assignment.
5. **Comparing Revisions to Submission Timeline:**
    - Cross-reference the revision timestamps with the submission timeline provided by the submitter.
    - Check if the revision timestamps align with the claimed timeline of document creation and edits.
6. **Inconsistencies and Anomalies:**
    - Look for any unusual patterns or discrepancies in the edit times, revision counts, or timestamps.
    - Anomalies may indicate potential issues with the authenticity of the work.

By examining the file properties and gathering information on edit time, revisions, timestamps, and authorship, teachers can assess whether the claimed editing history aligns with the expected creation and editing process of the work. This approach helps in evaluating the authenticity and integrity of the submitted piece.

The *HSC: All My Own Work* program as Guidelines for The Use of GAI In Assessment Tasks

Completing Assessment Tasks Honestly

HSC: All My Own Work is a program designed to help HSC students follow the principles and practices of good scholarship. This includes understanding, valuing and using ethical practices when locating and using information as part of their HSC studies.

Students who have completed the program will also know about penalties for cheating and how to avoid malpractice when preparing their work for assessment.

Teachers and students should use the knowledge and skills gained in the process of the *HSC: All My Own Work* program to cultivate ethical research, citation, and originality in the completion of their Stage 6 Assessment Tasks. As the program quips Year 10 students with essential principles of academic integrity and responsible research, it provides a solid foundation for ethical use of AI in assessment tasks in the senior school

At the conclusion of the *HSC: All My Own Work* program, students will understand:

- How academic integrity and responsible research practices translate to giving appropriate credit when utilizing AI-generated content or tools.
- How to critically evaluate sources, promoting discernment of the reliability and credibility of AI-generated information.

- An awareness of plagiarism and ethical practices, ensuring all students use AI responsibly and refrain from submitting AI-generated work as their own.

By integrating the principles taught in the *HSC: All My Own Work* program, teachers and students can approach AI usage in their courses and assessment tasks ethically, preserving academic integrity and fostering responsible AI engagement for their academic growth.