

ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity

Sasha Nikolic, Scott Daniel, Rezwanul Haque, Marina Belkina, Ghulam M. Hassan, Sarah Grundy, Sarah Lyden, Peter Neal & Caz Sandison

To cite this article: Sasha Nikolic, Scott Daniel, Rezwanul Haque, Marina Belkina, Ghulam M. Hassan, Sarah Grundy, Sarah Lyden, Peter Neal & Caz Sandison (2023) ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity, *European Journal of Engineering Education*, 48:4, 559-614, DOI: [10.1080/03043797.2023.2213169](https://doi.org/10.1080/03043797.2023.2213169)

To link to this article: <https://doi.org/10.1080/03043797.2023.2213169>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 26 May 2023.



Submit your article to this journal [↗](#)



Article views: 8698



View related articles [↗](#)



View Crossmark data [↗](#)



ChatGPT versus engineering education assessment: a multidisciplinary and multi-institutional benchmarking and analysis of this generative artificial intelligence tool to investigate assessment integrity

Sasha Nikolic ^a, Scott Daniel ^b, Rezwanul Haque ^c, Marina Belkina^d, Ghulam M. Hassan ^e, Sarah Grundy^f, Sarah Lyden ^g, Peter Neal ^f and Caz Sandison ^a

^aFaculty of Engineering and Information Sciences, University of Wollongong, Wollongong, Australia; ^bSchool of Professional Practice & Leadership, University of Technology Sydney, Sydney, Australia; ^cSchool of Science, Technology and Engineering, University of the Sunshine Coast, Sippy Downs, Australia; ^dCollege, Western Sydney University, Sydney, Australia; ^eSchool of Computer Science & Software Engineering, University of Western Australia, Perth, Australia; ^fSchool of Chemical Engineering, University of New South Wales, Sydney, Australia; ^gSchool of Engineering, University of Tasmania, Hobart, Australia

ABSTRACT

ChatGPT, a sophisticated online chatbot, sent shockwaves through many sectors once reports filtered through that it could pass exams. In higher education, it has raised many questions about the authenticity of assessment and challenges in detecting plagiarism. Amongst the resulting frenetic hubbub, hints of potential opportunities in how ChatGPT could support learning and the development of critical thinking have also emerged. In this paper, we examine how ChatGPT may affect assessment in engineering education by exploring ChatGPT responses to existing assessment prompts from ten subjects across seven Australian universities. We explore the strengths and weaknesses of current assessment practice and discuss opportunities on how ChatGPT can be used to facilitate learning. As artificial intelligence is rapidly improving, this analysis sets a benchmark for ChatGPT's performance as of early 2023 in responding to engineering education assessment prompts. ChatGPT did pass some subjects and excelled with some assessment types. Findings suggest that changes in current practice are needed, as typically with little modification to the input prompts, ChatGPT could generate passable responses to many of the assessments, and it is only going to get better as future versions are trained on larger data sets.

ARTICLE HISTORY

Received 28 March 2023

Accepted 8 May 2023

KEYWORDS

Artificial intelligence (AI);
assessment; ChatGPT;
engineering education;
GPT-3; integrity

Introduction

At the start of 2023, the authors of this paper, hailing from seven different Australian universities, were bombarded with information related to ChatGPT, an artificial intelligence (AI) powered, web-based chatbot. This included emails from university leaders such as Vice Chancellors, Deputy-Vice Chancellors, Deans, and Heads of School, aimed at raising awareness about the software. Invitations were also extended to attend information sessions to gain insights into the capabilities of ChatGPT, but such sessions contained very little on the known impact on current assessment practices. Teaching and learning departments were instructed to develop resources on AI software to educate staff

CONTACT Sasha Nikolic sasha@uow.edu.au

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

quickly, shared through university pages, such as Lei (2023b), and social media. With little empirical evidence available, only guiding principles were made available. Additionally, concerns were raised about the integrity of assessment practices, with reports suggesting that ChatGPT could pass some assessments (Gilson et al. 2022) and evade plagiarism detection by humans or software (Else 2023). Despite these concerns, the rapid pace of its rollout meant there was a dearth of relevant education research, especially in engineering education. The widespread hype, confusion, and duplication of workload across institutions underscored the need to understand the potential impact of ChatGPT on the integrity of engineering education assessment. Therefore, this study aims to answer the research question *'How might ChatGPT affect engineering education assessment methods, and how might it be used to facilitate learning?'*

As outlined in the literature review, the performance of chatbot technology is rapidly improving. For that reason, it needs to be acknowledged that the comparisons and analyses made herein could be out of date before this article is even published. Consequently, this article aims to establish a benchmark of ChatGPT's performance in a diverse range of assessment tasks during the first quarter of 2023. This benchmarking data will be invaluable in comparing changes over time as the technology continues to advance. The authors plan to repeat this study to test against future releases. In addition to benchmarking, the data analysis conducted in this study will identify areas of weakness and strength within the scope of engineering education assessment. The study will also explore potential adaptations to the new landscape brought on by powerful artificial intelligence technology and consider options to embrace it. GPT-4 was released days before the submission of this study. Some commentary regarding early claims of its performance is included.

Literature review

A chatbot is a computer program designed to engage in conversations with humans and provide solutions to their questions (Dahiya 2017). Chatbot technology is not new, with the first known program called ELIZA being developed in 1966, using simple pattern-matching techniques together with a template-based response mechanism (Adamopoulou and Moussiades 2020). Chatbot technology evolved, and a major breakthrough in the technology was made in 1995 with the award-winning program ALICE, combining pattern-matching with artificial intelligence (the ability of machines or computer systems to perform tasks that normally require human intelligence) to provide a natural language user experience (Bani and Singh 2017). The more natural the communication experience, the more likely the user would feel that they are having a real conversation and not a simulated one. Natural Language Processing (NLP) is a subfield of artificial intelligence providing the scaffold that allows chatbot programs to understand one or more human languages (Khanna et al. 2015). One of the next major advancements in the field came through virtual personal assistants like Apple's Siri in 2011 and Amazon's Alexa in 2014 (Adamopoulou and Moussiades 2020). These are technologies known by name across much of the world.

OpenAI's ChatGPT (officially Chat Generative Pre-Trained Transformer) released its popular GPT-3 version in October 2020, following the release of GPT-2 in February 2019 and GPT-1 in 2018. ChatGPT is a Large Language Model (LLM) that uses a form of NLP called 'unsupervised learning' to generate its responses. This involves training the model on large amounts of text data to learn patterns and relationships between words and phrases. When presented with a new prompt or question, ChatGPT uses its learned knowledge to generate a response that is contextually relevant and grammatically correct (OpenAI 2023b; Bubeck et al. 2023). The first model was based on 117 million parameters, the second on 1.5 billion parameters, and the third version (used in this study) on 175 billion parameters (OpenAI 2023c). As can be seen, the increase in training parameters in such a short time has been substantial. The size of training parameters is important because the software uses machine learning to autonomously learn (van Dis et al. 2023). With the increase in training size, GPT-3 can now capture even more complex patterns and relationships in language, resulting in more sophisticated and nuanced responses.

A famous measure of artificial intelligence is the Turing Test. Named after the pre-eminent computer scientist Alan Turing, it tests whether a human subject can distinguish which of two conversational partners is a fellow human being, and which is a machine. In two recent studies in a medical context, ChatGPT's responses were only weakly distinguishable from those generated by a real person (Hulman et al. 2023; Nov, Singh, and Mann 2023).

While ChatGPT has received considerable media attention in the chatbot space (Else 2023), it is not the only program of its kind. Other similar LLM's include Google's products Bert, Meena, XLNet, and Bard; Microsoft's Xiaolce and Bing ChatGPT, and Facebook's Blender (Agomuoh 2023). These models are capable of engaging in natural, human-like conversations and can be used for a wide range of applications, including customer service, education, and entertainment. Nevertheless, ChatGPT has been the most widely discussed language model thus far (Frieder et al. 2023), and has set a new record as the fastest growing consumer application in history (Hu 2023), making it the primary focus of this study. In terms of AI tools more generally, there are over 1300 available (Futurepedia 2023) at the time of writing, with new applications frequently being developed.

To obtain highly relevant output with large language models like ChatGPT, the user needs to carefully consider what the best input parameters are. For example, if a user provided an input such as 'what was the most popular car in 2020?' they would get a very different answer if the input was 'what is the most popular electric sedan in Australia in 2020?' As a result, the term *prompt engineering* is now commonly used to describe the process of modifying the input prompts to tailor the output and ensure specific qualities (and quantities) of generated output (White et al. 2023). The starting point for a prompt includes providing instruction, context, input data and output indicator (Lei 2023a). This has opened up a new, hot field of research and prompt engineers can attract six-figure salaries (Popli 2023).

Interestingly, other applications such as Wolfram Alpha (released in 2009), have been providing advanced mathematical capabilities that are well suited for engineering education for quite some time (Necesar and Pospíšil 2012). At present, ChatGPT's capability in mathematics, in contrast to its natural language capabilities, is claimed to be rather poor, falling short of the level of an average mathematics graduate student (Frieder et al. 2023). Therefore, this provides justification to explore in this study if ChatGPT is indeed any match for the numerical questions typically found in engineering education assessments. Within the scope of engineering mathematics, software like Wolfram Alpha have been used by some educators, as discussed in Necesar and Pospíšil (2012) and Kanwal (2018), as a learning aid rather than something to be feared or avoided as a threat to authentic learning and assessment. For this to become more widespread, we need to understand the technology in an educational perspective, we need to provide necessary safeguards, and update policies (Murugesan and Cherukuri 2023).

Concerns about ChatGPT potentially compromising academic integrity have arisen due to reports of its ability to pass various assessments and deceive experts. These include passing components of the medical licencing exam (Gilson et al. 2022); writing fake research abstracts that were not detected by experts (Else 2023); and passing law exams (Choi et al. 2023). The chatbot has been used and tested with varying success and accuracy with many use cases including writing computer programs, writing papers and essays, undertaking statistical analysis, summarising literature, and more (van Dis et al. 2023). In engineering education, there is a wide range of assessments used, some more frequently than others, and even before all the ChatGPT buzz, research indicated that we need to develop a greater understanding of assessment implementations such as how and why they are used (Nikolic et al. 2021; Hassan 2011). It has been suggested that AI technology can help educators with the design and implementation of assessment (Swiecki et al. 2022). Again, this provides justification of the need for this study to investigate the impact ChatGPT can have on engineering education assessments of various types.

Academic integrity fears have been heightened by the COVID-related shift of assessments in tertiary education to online formats which has seen increases in cheating behaviours, providing a pathway to encourage student use of ChatGPT or related technologies (Susnjak 2022). A systematic literature review of online exam cheating by Noorbehbahani, Mohammadi, and Aminazadeh (2022)

outlined a variety of motivations as to why students may decide to cheat. They highlight that the primary reason being that students would cheat if the reward outweighed the risk. Factors they identified that motivate students to cheat could be grouped as teacher, institutional, internal and environmental reasons. Overall, if a student feels disadvantaged, the environment is conducive of cheating, and the assessment design is one that makes cheating easy, students will be more tempted to cheat. Many of these risk factors can be easily correlated to online or other non-invigilated assessment contexts. However, Noorbehhahani, Mohammadi, and Aminazadeh (2022) outline that cheating can be deterred using an anti-cheating atmosphere. They contend that ethical behaviour of staff, supportive learning environments, assessments designs that are difficult to cheat with and are complemented with cheating-detection tools, all supported by strong institutional rules and policies, help deter cheating. Deterring cheating is important, due to the flow-on risk of students continuing such unethical conduct in to their professional practice and the compounding problems associated with this (Ellis et al. 2020).

While ChatGPT has raised substantial integrity concerns, cheating is not a new phenomenon. For example, contract cheating has received much attention in recent years (Ellis et al. 2020). Plagiarism detection software, such as Turnitin, has helped uphold and support the communication of the importance of academic integrity at tertiary institutions for some time (Alua, Asiedu, and Bumbie-Chi 2023). Beyond written assessment, proctoring software (i.e. software to support the online supervision of live assessments) has been used to support online exams, but has faced many ethical concerns and is more a deterrent rather than a tool to stop cheating (Alin, Arendt, and Gurell 2023). At the time of writing, ChatGPT can overcome plagiarism tools like Turnitin because it produces a unique output for each query, even if the query is repeated. To overcome this, a new breed of software tools and extensions like GPTZero (2023) are becoming available to help predict if a document was written by a large language model. However, they are in their infancy and need substantial refinement to help better support academic integrity.

The literature outlined above has suggested that proctoring software and traditional and future detection techniques currently provide limited value in upholding academic integrity. In an environment where ChatGPT can provide unique output and can provide passing-level answers to some assessment tasks, some re-evaluation of current practice is needed. As was highlighted by Noorbehhahani, Mohammadi, and Aminazadeh (2022), strong institutional rules and policies, together with assessment designs that deter cheating, are needed. Therefore, we must reflect on current assessment practices and determine what is the new best practice, which is what we are seeking to accomplish in this study.

While integrity is one concern, as engineers, embracing new technology to advance the way we do things is a must. Therefore, we need to work through the weaknesses and determine how artificial intelligence tools like ChatGPT can be used to advance engineering education. The benefits of integrating such technology into the classroom (rather than fearing it) is demonstrated by Necesal and Pospíšil (2012). For this reason, in the study where possible, we try and explore how ChatGPT can aid the learning process through undertaking the assessment tasks. When ChatGPT was asked how it could benefit engineering education, the following points was the direct output (OpenAI 2023e):

1. Improved Engagement and Student Experience

One of the key benefits of ChatGPT in engineering education is its ability to enhance student engagement and improve the overall student experience. ChatGPT can interact with students in a conversational manner, providing them with personalized feedback and allowing them to ask questions in real time. This type of interaction can help students feel more connected to the material and can lead to a deeper understanding of the subject concepts. Additionally, the conversational style of ChatGPT can make the learning experience more engaging and enjoyable for students, reducing the boredom and disinterest that can arise in traditional lecture-style teaching.

2. Increased Accessibility and Inclusivity

Another important benefit of ChatGPT in engineering education is its ability to increase accessibility and inclusivity. ChatGPT can provide students with instant feedback and support, regardless of their location or time zone. This can be particularly useful for students who are unable to attend traditional in-person classes or who require additional

support to fully grasp the subject material. Additionally, ChatGPT can be used to provide real-time translation of subject material, making engineering education more accessible to students who are non-native English speakers.

3. Enhanced Assessment and Feedback

ChatGPT has the potential to revolutionize the way engineering students are assessed and receive feedback on their performance. ChatGPT can be used to provide students with instant feedback on their assignments and assessments, allowing them to quickly identify areas where they need to improve. This can be especially beneficial for students who are unable to attend regular office hours or who need additional support outside of class time. Additionally, ChatGPT can be used to grade assignments and provide feedback, freeing up valuable time for instructors who would otherwise be spending hours grading papers.

4. Improved Learning Outcomes

The use of ChatGPT in engineering education has the potential to improve learning outcomes for students. ChatGPT can provide students with instant feedback on their progress and help them to identify areas where they need to improve. Additionally, the conversational style of ChatGPT can make the learning experience more enjoyable and engaging, leading to a deeper understanding of the subject concepts. Furthermore, ChatGPT can provide students with personalized learning experiences, tailoring the learning process to their individual needs and preferences.

When ChatGPT was asked about its limitations regarding engineering education, the following points was the direct output (OpenAI 2023d):

1. Training Data Bias

As a machine learning model, ChatGPT has been trained on a large corpus of data, and any biases present in that data will be reflected in the output generated by the model. For example, if the training data contains stereotypical gender roles or perpetuates cultural or racial biases, ChatGPT's output may reflect these biases as well.

2. Inaccurate Information

Despite its advanced language generation capabilities, ChatGPT is not always able to provide accurate information. This is due to the limitations of the training data and the fact that ChatGPT is not capable of independent judgement. As a result, it is important for users of ChatGPT to verify the information it provides before using it in engineering education.

3. Limitations of AI

ChatGPT is not capable of independent judgement or decision-making and relies on the input provided to it to generate its output. This means that the information provided by ChatGPT may be limited or incorrect, and that it is important for instructors and students to use critical thinking and judgement when interpreting its output.

As can be seen above, the English-level capabilities of the software can be rather good. This capability has led to contentious practices in using ChatGPT to co-author research articles (Stokel-Walker 2023). ChatGPT raised some valid points as to why the technology should not be feared but embraced within engineering education. The limitations raised provide an opportunity to reflect and find opportunities to integrate these weaknesses into different assessment tasks. This could provide a pathway to new opportunities to integrate ChatGPT and similar tools into classroom learning activities.

Research activity

Nine authors from seven different Australian universities, all with different engineering backgrounds, came together to tackle the common problem identified and highlighted in the introduction and literature review. That is, there was great institutional demand to develop a ChatGPT-based understanding of the weaknesses of current assessment practices, as well as identifying strengths and new opportunities. Therefore, a process was developed to answer the stated research question *how might ChatGPT affect engineering education assessment methods, and how might it be used to facilitate learning?*

To accomplish this, the initial phase consisted of the authors exploring the capabilities of ChatGPT in responding to a range of assessment tasks and determining what information in the assessment prompt was important. This was completed using the publicly available webpage at <https://chat.openai.com/chat>, a free version of ChatGPT.

Team meetings followed and were used to refine the investigation and data collection process. From this, procedures and templates were formulated to create consistency across each team member. A decision on which subjects would be included was made to ensure as much diversity in subject matter and assessment implementation as possible. Investigation and reporting were cascaded allowing each member to report their results to the team, scaffolding and sharing implementation to strengthen alignment. The first author played a gatekeeper role, ensuring consistency and best practice in 'prompt engineering'. This was to ensure that prompts were being applied to give ChatGPT every chance to pass the assessment tasks. The first author is also a Head of Students for undergraduate engineering, a role that requires having oversight into the structure for eleven different engineering majors. This experience provided the skills necessary to reflect on the individual efforts of each team member.

The ultimate aim was to determine if a student could use ChatGPT to pass the different assessment tasks and how difficult that may be. This was the driving force behind the research methodology. Assigning a pass or fail was the primary recording method. For questions that had a black or white answer (such as marks for numerical questions with only one correct answer), an actual grade was recorded. For other assessments, a grade beyond pass or fail was avoided due to the bias in knowing that the output was AI-generated. Each team member would need to determine if, at a minimum, a passing grade could be reached. The advantage of this approach was that each team member could further adapt the input of the questions provided to ChatGPT to determine if there was a way to produce a passable output. Hence the importance of the gatekeeper role mentioned above. This would allow team members to best report on the impacts of input modification (that is, changes to ChatGPT's prompt). The reflective process of interacting with changes to the input and output, created a connection to the second component of the research question (*how might it be used to facilitate learning?*). It was discovered during the investigation stage, that as the team uncovered what ChatGPT could or could not do, reflectively the team was also identifying opportunities.

For this reason, the team classified the level of input modification required to obtain a pass. The classifications are as follows:

N/A: related to modification not being applicable (e.g. ChatGPT could not take the place of a student for an oral presentation)

Fail: if a modification was applied, the output resulted in work that would fail.

None: correlated to an output that would give a passing grade simply by copying the assessment task information/question directly into ChatGPT.

Minor: related to some minor modification being needed to the input in order to get a pass grade. These generally resulted in the need to input specific parameters or instructions that provided better guidance of the structure and content required in the output. For example, simply rewording the question to use the prompt engineering structure instruction, context, input data and output indicator (Lei 2023a).

Major: correlated to substantial modifications required to achieve a passing grade. In essence this would require a student that has a solid understanding of what the output needs to be and has the technical understanding to experiment with the input to achieve the required result.

The team agreed to use shared procedures and templates to ensure consistent documentation. Examples of the input and output were required to be recorded to allow team analysis of the modification levels applied to the input. At the end of the data collection period, the team reconnected to analyse the results, share their findings and undertake analysis for this paper.

The data collection process described above was formally conducted as follows:

- For each subject, the team member responsible needed to test all assessment tasks outlined on the subject outline (the formal university documentation regarding subject structure).

- The team member would first attempt to copy and paste the question into ChatGPT.
 - If the output produced an output that would pass, they would move on to the next question.
 - If the output would not produce a pass grade, the team member would reflect on the relationship between the prompt and output to reengineer the prompt. This process was aided by the structure outlined above.
 - If the output would result in a pass, they would move to the next question, or if the reflection warranted it, further prompt engineering exploration was undertaken to explore what it would take to get a better result to aid the discussion.
 - If the output did not result in a pass, they would continue to reengineer the prompt. When in doubt, the gatekeeper was consulted. This process continued until it was determined that a passing grade was not possible, or a passing grade was achieved.
- When all assessments were completed, the results and examples were provided to the team. This allowed for feedback and knowledge transfer on practical prompt engineering approaches. The next team member would then repeat the process on their subject.
- If the assessment task involved a random question set, the team member just needed to complete the assessment task once. That is, they did not retake the test to check different combinations of questions.

In terms of analysis and presentation, the team considered what data academics would want to see and how they would read the paper. From recent information sessions on ChatGPT at authors' universities, two themes emerged. Academics wanted to know practical information on the impact on the subject they taught, and they wanted to know which assessments were most vulnerable and which were not. With this information we believe many readers will read the impact on the subject they teach (or closest to it) and then read the information about the assessment types. For this reason, data analysis in the following section is separated into a subject-by-subject and an assessment-type approach.

The assessment types were bucketed into categories of best fit by the team after an analysis of the requirements of each assessment type conducted in each subject. The categories were defined as:

- **Online Quizzes:** tasks that used an online quiz format using an e-learning platform.
- **Numerical (Assignments and Exams):** assessments tasks where the answers are numerical in nature (e.g. calculation based) and are completed in a written format (not online).
- **Code Submission:** assessments requiring the submission of programming code.
- **Oral:** assessments comprising presentations, interviews, pitches and quality participation in discussion.
- **Visual:** visual documents (e.g. mind map) and evidence (e.g. completion certificate)
- **Written (Experimentation-based):** written activity associated with experimentation or laboratory work.
- **Written (Project-based):** written assessment activity associated with project work (e.g. project report)
- **Written (Reflective & Critical Thinking-based):** written assessment tasks that focused on reflective and critical thinking (e.g. reflection on student experience, strengths and weaknesses)
- **Written (Research-based):** assessments focused on research-based writing (e.g. thesis).

Just before paper submission, the new version of ChatGPT (module 4) was released via paid premium subscription. While access to the full feature set was not readily accessible, we managed to check against some of the ChatGPT limitations identified in this research. This update is provided in a commentary section.

A diverse range of both undergraduate and postgraduate subjects from across the seven universities were selected to provide a diverse picture of the impact ChatGPT has on engineering education

assessment. The subjects were also selected for both their similarity and differences in assessment types. Subject codes and location are masked to respect the sensitivity of the data. Although different terms are used in different universities and contexts, in this study we use the term 'subject' to refer to a single unit of study equivalent to about one-quarter of a full-time semester load, requiring roughly 100 h of student work. Subjects investigated include:

First-Year Foundational:

- **Engineering Physics:** a first-year undergraduate subject for all engineering disciplines. The subject serves as an introduction to the fundamentals of engineering physics with appropriate applications in a wide range of engineering and industrial design systems.
- **Maths:** a first-year undergraduate subject common for all engineering disciplines focused on Calculus and Linear Algebra. A contextual focus from various engineering disciplines is used.

Technical:

- **Introductory Programming:** a first-year undergraduate subject for all engineering disciplines and other related fields. Using Python, students develop computational thinking to solve problems with a focus on building small programs for specialised tasks.
- **Manufacturing Technology:** a second-year undergraduate subject with a primary focus for mechanical engineering. Individually and through teamwork students develop understanding of modern manufacturing processes, production systems and quality management systems.
- **Engineering Laboratory:** a third-year undergraduate laboratory subject tailored for chemical engineers. In teams, students are presented with open-ended projects requiring them to lead, plan and execute laboratory work.
- **Sustainable Product Engineering and Design:** an undergraduate first-year to final-year, multi-disciplinary engineering and science subject. Issues with respect to the sustainability, safety, engineering profession, laboratory, pilot plant are explored. Students also gain skills in information literacy.
- **Renewable Energy and Electrical Power:** a fourth year undergraduate and co-badged postgraduate subject for electrical engineers. Students develop knowledge in renewable energy technologies, energy storage systems, power electronics interfaces and associated control, and renewable energy system design and implementation.

Social context of engineering/professional skills:

- **Sustainable, Environmental & Social Impacts of Technology:** A senior-level core undergraduate subject looking at the social dimensions of emergent technologies. Students learn about ethics, engagement and consultation with stakeholders, public policy, sustainability, and other contextual considerations, and then apply these concepts to exploring an emergent technology, such as self-driving cars or facial recognition software, from a range of perspectives.
- **Workplace Practice & Communication:** a multi-disciplinary postgraduate work-integrated learning (WIL) subject. This subject is focused on developing key employability skills required to be successful in the Australian workforce. Students refine their professional communication skills and work in a team as consultants working on real industry problems.

Research:

- **Engineering Research:** co-taught between fourth year undergraduate and postgraduate students (postgraduate assessment structure used in the study). The focus is on the

development of skills in framing a research problem, developing a research design, design of data collection analysis and interpretation frameworks.

Limitations:

There are several potential limitations with this study. First and foremost is the dynamically changing landscape of AI-based chatbot technology, with for example the release of GPT-4 a few days before submission. This means that by the time of publication, the ability of ChatGPT to respond to assessment task prompts will likely far exceed that reported here. However, we see this study as an opportunity to benchmark the performance of ChatGPT with regard to engineering education assessment, at this particular moment in time in early 2023.

All assessment tasks for all subjects are treated as if they could be completed via a technology that would allow access to ChatGPT. Some of the assessment tasks considered were designed to be conducted via an in-person format. Therefore, difficulty levels could change if designed for a different mode. For example, the difficulty setting of an open book test could be different to that designed for a closed book test. Regardless of this limitation, a benchmarking of current practice is observed.

Undergraduate engineering students in Australia typically complete something like 30 or more stand-alone subjects, covering the breadth and depth of skills and knowledge required for graduates. In this study, we have analysed ChatGPT's responses to assessments from around 10 subjects. One possible limitation, therefore, is the representativeness of our subject coverage. However, we feel we have pre-empted this concern by evaluating subjects from across the engineering curriculum, including first year foundational (maths and physics that underpins engineering content), technical, research, and professional skills from a range of year levels and from both common core and discipline-specific streams. While not exhaustive, we feel we have included sufficient diversity to give an accurate sense of ChatGPT's ability in addressing contemporary engineering education assessments.

While this study is focussed on ChatGPT, there are other artificial intelligence tools, some that build upon ChatGPT, that may be better suited to certain assessment tasks. Therefore, readers need to be aware of this self-imposed boundary. Using other AI tools may lead to different results.

A final potential limitation is the variability in assessment practices and rubrics in different institutions, and in our own different biases and interpretations of ChatGPT responses given our awareness they are generated by artificial intelligence. Another aspect of this is that in tweaking the assessment prompts to improve the ChatGPT response, we have a strong pre-conceived sense of quality as it is we ourselves that are evaluating and assessing student work, giving us better insights into how to adjust the prompt to improve the outcome. We have addressed these potential variabilities by using a consistent approach in evaluating ChatGPT responses as pass/fail and a straightforward framework to characterise the extent to which assessment prompts have been tuned to achieve improved outcomes.

Results and discussion

The results and discussion are separated into four different sections:

- *Subject Overview & Benchmarking*: A summary of the results and experience at a subject-by-subject level
- *Assessment Integrity*: An analysis by assessment type, outlining weaknesses and strengths and reflecting on the results to consider how ChatGPT can be integrated as a tool to aid learning.
- *Recommendations*: Steps researchers can take to build upon the findings from this study
- *ChatGPT-4 Commentary*: Some basic analysis and commentary on the recent release

The two data analysis sections have been undertaken using the framework of a SWOT (Strengths, Weaknesses, Opportunities and Threats). SWOT is a popular framework used in academic peer-reviewed literature and used by consultants, trainers and educators around the world (Helms and Nixon 2010). The analysis is centred on Strengths, Weaknesses and Opportunities as the threat is well defined as being ChatGPT. However, future threats are further outlined in the section ChatGPT-4 Commentary.

Subject overview & benchmarking

The first component of the research study consists of a benchmarking activity to set a baseline of comparison for measuring the advancement of ChatGPT into the future, and also allowing others undertaking similar subjects to compare. This data-intensive information is provided in Appendix A. A subject-by-subject reflection is provided to summarise the impact ChatGPT had on a student's ability to pass the subject. The team plans to reproduce the study against future versions of ChatGPT to monitor progress. The authors would hope that this study encourages other researchers to replicate the process and compare results across subject and assessment type.

Note that the following analysis is predicated on the assumption that the assessment tasks could be completed in an online environment wherever applicable. That is, if a test was conducted face-to-face in a subject, for analysis we assumed the test was conducted online and the student had the ability to access ChatGPT. The only exceptions to this were lab-based practical tasks.

Physics

Subject Summary: The subject applies the laws and principles of engineering physics to areas such as force systems and equilibrium, dynamics, motion, vibration and wave phenomena, and electricity and magnetism.

Assessment Types: Closed book exams, both in-session and end of session make up 75% of a student's grade. The other 25% is allocated to laboratory work with the assessment comprising of a Laboratory report marked at the end of the class.

Result: In order to pass the subject, students would be required to manipulate the input for many questions. In doing so, an overall mark of 56.9% could be achieved, a pass grade.

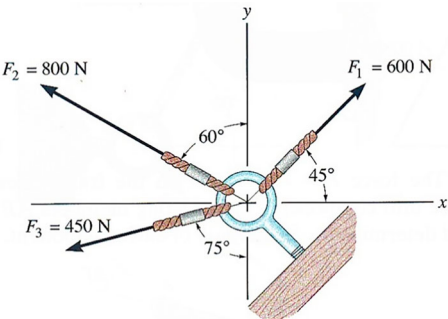
Integrity Strength: Questions containing figures and tables require the student to have some basic understanding of how to translate the question together with figure and/or table into words that ChatGPT could understand. For example, Table 1 below shows the descriptive text required for input in order for ChatGPT to provide the correct answer. For a student feeling the need to cheat, such an input may not seem obvious and could be time consuming. Student confusion can be amplified when considering the different symbols of the physics quantities, and equations provided in the text format that need to be rewritten in the correct mathematical style.

For the questions with tables, ChatGPT also can be used. Data from the table needs to be typed into the question. Adding the instruction 'present answers in a table', ChatGPT was able to present solution in a tabular format. However, at most only half of the columns were calculated correctly.

Integrity Weakness: ChatGPT provided the best solutions for simple questions, with no diagrams, tables, or multiple steps. Simple questions could obtain the correct result with no modification and provided reasonable explanation.

Opportunity: Having face-to-face assessment and a traditional laboratory supports academic integrity. If face-to-face assessments were retained, ChatGPT could be used to support student learning. ChatGPT provides great annotations and students can be encouraged to compare their solution with ChatGPT and use it for revision. Similarly, and while the success rate of ChatGPT is temporarily low, it can be used for 'flipped assessments'. Students can be asked to solve the problem with ChatGPT and then mark the output.

Table 1. Example of translating a figure into ChatGPT (Image from (Merriam and Kraige 1993)).

Original question	Question modification in ChatGPT
<p>The screw eye shown below is subjected to three forces F_1, F_2 and F_3.</p>  <p>Calculate the x and y components for each of the three forces F_1, F_2 and F_3 and their resultant R by Transferring to your written paper and completing the table below showing your calculations</p>	<p>The screw eye is subjected to three forces $F_1 = 600$ N East 45 degrees North; $F_2 = 800$ N North 60 degrees west; $F_3 = 450$ N South 75 degrees West. Calculate the x and y components for each of the three forces F_1, F_2 and F_3. Use calculated vector components to find the magnitude and direction of resultant force</p>

Maths

Subject Summary: The subject consists of two strands, Calculus and Linear Algebra. The Calculus strand covers differential calculus and introduces integral calculus. The Linear Algebra strand covers matrices, determinants and applications of these in the sub-topic of vector geometry.

Assessment Types: Online quizzes to reinforce lecture content comprised 10%, three written homework assignments totalled 20%, a mid-session exam covering content from the first half of semester contributed 35% and a final exam covering content from the second half of semester made up the remaining 35%. Both exams had a mixture of true/false, multiple choice and short-answer (working required) questions.

Result: To pass the subject, students would be required to learn the syntax for the mathematical expressions and manipulate the input for many questions. In doing so, the maximum overall mark achieved could be 58%, a pass grade. The final answer for many of the questions was wrong, but partial marks given for the correct components would allow students to pass the subject.

As Wolfram Alpha has been the ChatGPT of the mathematical world for some time, a comparison to this more established platform was also made.

Integrity Strength: ChatGPT does not calculate very well algebraically nor arithmetically, often resulting in a different answer each time it is asked to regenerate the response. If given a simple arithmetic question, ChatGPT can calculate correctly. However, when the arithmetic or algebra is embedded in a longer question, ChatGPT fails more often than it succeeds.

Integrity Weakness: ChatGPT gives very good instructions on the processes involved in answering questions. ChatGPT can also answer questions that have traditionally not been possible with online platforms, such as Wolfram Alpha, like ‘what is the best substitution to use to evaluate this integral?’. It also did very well with True/False questions. Interestingly, if some questions were repeated a wrong answer could be generated after the correct one was given.

A student who uses ChatGPT for the first assignment may realise quickly that ChatGPT is not reliable for calculation questions, but ChatGPT can tell the student what to do to answer the question, and more often than not, the instructions are correct and reliable. The student could then go to Wolfram Alpha for the correct calculation. Below is an example of how ChatGPT can give the correct instruction but make mistakes in calculation.

Entered Question:

Find the eigenvalues and eigenvectors for the matrix $A = \langle 1, 1, 2; 0, 2, 3; 0, 0, 3 \rangle$.

ChatGPT Response:

To find the eigenvalues and eigenvectors of matrix A , we need to solve the characteristic equation $\det(A - \lambda I) = 0$, where I is the identity matrix of the same size as A and λ is the eigenvalue.

First, let's write out the matrix A and the corresponding characteristic equation:

$$A = \langle 1, 1, 2; 0, 2, 3; 0, 0, 3 \rangle$$

$$\det(A - \lambda I) = \det(\langle 1-\lambda, 1, 2; 0, 2-\lambda, 3; 0, 0, 3-\lambda \rangle) = 0$$

Next, we expand the determinant along the first row to get:

$$(1-\lambda) [\det(\langle 2-\lambda, 3; 0, 3-\lambda \rangle) - 3\det(\langle 0, 2-\lambda; 0, 3-\lambda \rangle)] - 1[0 - 2\det(\langle 0, 2-\lambda; 0, 3-\lambda \rangle)] + 2[0 - 0] = 0$$

Simplifying and rearranging the terms, we get:

$$(1-\lambda)(\lambda^2 - 6\lambda + 9) - 2(\lambda-3)^2 = 0$$

Expanding and collecting like terms, we get:

$$\lambda^3 - 6\lambda^2 + 9\lambda - \lambda^3 + 6\lambda^2 - 9 - 2\lambda^2 + 12\lambda - 18 = 0$$

Simplifying, we get:

$$\lambda^3 - 9\lambda^2 + 27\lambda - 27 = 0$$

This is a cubic equation, which can be solved using various methods. However, we can notice that $\lambda = 3$ is a root of the equation, since substituting $\lambda = 3$ gives:

Analysis: The instructions given by ChatGPT are correct. Indeed, the solution continues on to find the eigenvectors, and the method used is correct. However, the algebra and arithmetic are incorrect. There are three eigenvalues, but with the mistakes in finding the characteristic equation, ChatGPT only found one. On another iteration of this question, it found two. On yet another iteration, it found three. In all cases, the eigenvectors were incorrect.

ChatGPT can also be used for the more theoretical questions and questions that require deduction, whereas Wolfram Alpha is only for calculations. From this, a prediction can be made that if the strengths of the two platforms were merged, most mathematical questions could be answered correctly with appropriate instruction provided and with little student effort. See the section ChatGPT-4 Commentary for important updates regarding this.

Opportunity: Given the quality of the theoretical answers provided by ChatGPT, students could use it as a personal tutor to help them study. For example, a student struggling to find eigenvalues and eigenvectors, could use ChatGPT to outline the procedure and explain the various steps. A student could continue to ask theoretical questions until they are satisfied they understand. This kind of help would be available continuously and students can ask questions without fearing their question may be considered 'a dumb question'.

Introductory programming

Subject Summary: The purpose of this subject is to develop programming as well as computational thinking skills in engineers. Range of problems are solved using fundamentals of Python programming language including loops, conditions, lists, dictionaries, recursion and file handling. Additionally, computational thinking skills which include problem decomposition, abstraction, pattern recognition and algorithms are used to design the solutions.

Assessment Types: Five online labs comprised 10% to reinforce lecture contents and simple programming skills, two large complex projects totalled 35% to cover computational thinking and complex level of programming skills, and a final exam of 55% covering range of problems to evaluate programming fundamentals and computational thinking skills.

Result: Introductory-level programming is a key strength of ChatGPT. If all assessments were online and students had access to use ChatGPT, an overall mark of 68.2% could be achieved, a comfortable pass (credit grade given).

Integrity strength: While such tasks achieved a high success rate, ChatGPT did struggle with more complex tasks. Within the subject, a major project involved writing a program to read a CSV file and analysing the data to provide statistical outputs. While many attempts were made to get ChatGPT to undertake this task, the requirements for the solution were beyond its capabilities.

Integrity weakness: ChatGPT did an outstanding job of providing correct output to simple coding requests and translating worded instructions into code. It provided good explanations to support the understanding of the code generated. An example of its ability to understand instruction, provide code and explain its design is presented:

Entered Question:

A psychology student is carrying out an experiment in which she repeatedly asks people to “think of an object, any object”. She wants to see what objects people think of under this situation. She enters each guessed object into a text file, one object per line.

Help her to analyse the data by writing a function `make_dictionary(filename)` that reads the named file and returns a dictionary mapping from object names to occurrence counts (the number of times the particular object was guessed). For example, given a file `mydata.txt` containing the following:

Dog, triceratops, persian cat, dog, persian cat, large white fluffy thing, persian cat

the function would return a dictionary-like

```
('dog': 2, 'persian cat': 3, 'triceratops': 1, 'large white fluffy thing': 1)
```

[The order of keys in a dictionary is arbitrary, so the objects might be in any order when the dictionary is printed.]

ChatGPT Response:

Here's one way to implement the `make_dictionary` function:

python

Copy code

```
def make_dictionary(filename):
    obj_dict = {}
    with open(filename) as f:
        for line in f:
            obj = line.strip()
            if obj:
                obj_dict[obj] = obj_dict.get(obj, 0) + 1
    return obj_dict
```

Here's how it works:

We initialize an empty dictionary `obj_dict` to store the object names and occurrence counts.

We open the file `filename` using a `with` block to ensure that the file is properly closed when we're done with it.

We iterate over each line in the file using a `for` loop. The `strip()` method is used to remove any leading or trailing white space from the line.

If the resulting object name is not empty (i.e. it has at least one non-space character), we update the occurrence count in `obj_dict` by 1. We use the `get()` method to retrieve the current occurrence count for the object name (if it exists) or 0 (if it doesn't exist).

After processing all lines in the file, we return `obj_dict`.

This implementation handles the case where the file is empty or contains only blank lines, since empty object names are ignored.

Opportunity: ChatGPT provided a very good explanation to simple programming problems which can help students to strengthen their fundamental programming concepts. ChatGPT can act as a tutor available at all times. For large complex tasks, one method could be to use computational thinking skill of decomposing the problem into small parts to get a solution for each part from ChatGPT and combine them. However, if a student could get it to work, then they would have had the ability to complete the assessment themselves anyway. While this activity was suitable as a major project, something this complex would not be suitable for an exam situation.

Manufacturing technology:

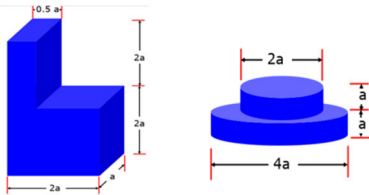
Subject summary: This course introduces the knowledge and skills of modern manufacturing processes, production systems and quality management practices to turn a conceptual idea into a globally competitive the finished product. The theoretical knowledge is reinforced with practical work, demonstrations and factory visits.

Assessment types: To assess the student's learning, a range of evaluation methods were employed, including written assignments, lab reports, oral conversations, and a centrally scheduled exam. The evaluation criteria were designed to ascertain the students' proficiency in various manufacturing processes and their analytical skills. The assessments aimed to measure the students' ability to demonstrate in-depth knowledge of diverse manufacturing techniques, analyse complex data sets, and interpret the results obtained from analytical solutions. The evaluations aimed to provide a comprehensive measure of the students' knowledge, skills, and abilities in the field of manufacturing processes.

Result: ChatGPT's success in this subject is contingent upon the type of questions presented. Due to the combination of questions used in the tested subject, ChatGPT received a marginal failure with a score of 47%. With a different assortment of questions, ChatGPT could have passed, highlighting the significance of assessment design and construction.

Integrity strength: The primary weakness was encountered when translating figures into words, which proved to be a challenging task. As a result, ChatGPT failed to provide the desired output in most cases. Additionally, table-based questions were challenging to translate into a useable input for ChatGPT to analyse. Given that numerous questions were based on figures and tables, this increased the failure rate. The subject comprised a hands-on face-to-face laboratory component, which meant that ChatGPT was unable to replace the student and perform the experiments on their behalf. The laboratory report was the basis for evaluating this task.

Table 2. Example of question difficulty levels for input into ChatGPT.

Question easily calculated by ChatGPT	Difficult question to translate in ChatGPT
A 15 mm diameter copper billet is extruded into a 10 mm diameter billet. Using the values, $\mu = 0.07$; semi die angle, $\alpha = 60^\circ$ and $s_o = 320$ MPa; calculate the extrusion force required. Also determine the die pressure.	Determine the shape factor of the following objects:
	

Integrity weakness: ChatGPT can solve numerical questions that can be directly inputted (simple copy and paste) and provides clear step-by-step explanations of the solution process. Traditionally, a method of generating unique student questions was to use the student's ID number to create distinct variables for calculation purposes. This was used to ensure no two students were solving the same equation. With ChatGPT, this uniqueness posed no major challenge to such problems as long as it had the necessary values to perform calculations. Table 2 demonstrates a question that ChatGPT can easily tackle, and another that is difficult to translate.

Opportunity: While ChatGPT could not generate the entire lab report, it could assist students with segments of the report. As long as they have the sufficient understanding of their experimental work (i.e. they should have already conducted the experiment, collected data, and analysed their findings) ChatGPT can assist with writing, formatting and technical aspects of their report, ultimately leading to better academic outcomes.

Engineering laboratory

Subject Summary: The focus is for students to develop and extend skills in designing and executing experimental investigations of engineering problems using small pilot-scale unit operations and analytical equipment. Students develop skills in analysis, critical thinking, communication, project management and teamwork.

Assessment Types: Written proposal/report work is the major focus area and is worth 66%. Supporting oral activities is allocated 29% with the remaining 5% is used for an introductory quiz. The presentations and reports comprising both technical and reflective communication skills.

Result: Due to the suggestive nature of most of the assessment tasks, no final grade estimation is provided, but indicatively it would be very difficult for a student to pass the subject using ChatGPT. In some cases, the responses (to some sections) were passable, but in general they lacked accurate and concrete examples from the actual experiments.

Integrity Strength: Substantial effort by a student would be needed to use ChatGPT to complete the assessment tasks to obtain reasonable marks. The student would firstly need to feed the system significant amounts of content:

- Assessment instructions,
- Marking rubric,
- Transcript of H5P pre-laboratory lesson, and
- Notes from the pre-laboratory briefing and from the conduct of the experiment.

This would be quite laborious, and it reaches a point where either:

- (a) It becomes more efficient for a student to simply write the document themselves, or
- (b) The student is simply using the system as a secretary to turn notes into prose.

The word limits with the current version of ChatGPT made using it more complex. This is because for some of the assessment tasks the responses required need to be longer than what ChatGPT will generate. Therefore, the proposal and report tasks were broken up into sections and asked in turn (e.g. 'Now, write the Methods section for the above experiment'). ChatGPT could also generate data tables for the experiment – some of which had no connection to the reality of the experiment.

This subject involves several reflection tasks (in reports and the final seminar). ChatGPT was able to employ various reflective schema (when prompted) and produced texts that would probably be passed by some staff when part of a lab report – but as mentioned above length of output was problematic. The output tended to be generic. The lack of specific examples was a clear sign that the 'author' had not done the experiment.

Integrity Weakness: The only time ChatGPT gave satisfactory responses without modifying the prompts was for the multiple-choice questions that form part of the introductory quiz that is largely formative in intention. Even here, the system got some questions wrong because they asked about information that was presented in person at the lab induction.

Opportunity: If ChatGPT was taught as a tool, combined with reflective and critical thinking exercises, ChatGPT could support student learning. This could be achieved, by providing a foundation that the student could work with to help improve their own writing and/or understanding of experimentation.

To put this into perspective an example of ChatGPT writing a Methods section is shown below:

Entered Question:

Write a method section for the above experiment. It should contain a concise description of the experimental procedures, a schematic of the apparatus, identification of the equipment and materials used, and descriptions of the sources of experiment errors and precautionary steps to minimise them (both systematic and random).

ChatGPT Output:

Method:

Experimental Procedures:

1. The packed column was set up for distillation by filling it with glass beads.
2. A temperature probe was inserted at the bottom of the column to measure the temperature of the distillate.
3. The distillation flask was charged with 100 mL of the 1 vol% ethanol feed.
4. The flask was then heated, and the distillate was collected in 5 mL increments until a total of 30 mL was collected.
5. The samples were labelled and stored for further analysis.

Apparatus:

1. Packed column (with glass beads).
2. Temperature probe.
3. Distillation flask.
4. Heating mantle.
5. 5 mL vials for sample collection.

6. Gas chromatograph (for ethanol analysis).

Materials:

1. 1 vol% ethanol feed solution.

Sources of Experiment Errors:

1. Errors in volume measurements: To minimize this error, all volume measurements were taken using calibrated measuring cylinders and pipettes.
2. Instrument errors: To minimize this error, all instruments were calibrated before the experiment.
3. Operator errors: To minimize this error, the operator followed the experimental procedures carefully and paid close attention to detail.

Precautionary Steps:

1. All glassware and instruments were cleaned before use.
2. All equipment was set up carefully to ensure that there were no leaks or other issues.
3. The experiment was performed in a well-ventilated area to prevent exposure to hazardous vapours.
4. Safety goggles and lab coats were worn throughout the experiment to prevent accidental exposure to hazardous substances.
5. All waste materials were disposed of according to laboratory safety protocols

What was generated is a good starting point but was not reflective of the setup used in the laboratory and therefore contained errors. In part this is because the prompt was rather basic. Critical thinking exercises could be used to help students through the thought process. A reflective process could help them make changes to the prompt to address the wrong assumptions made by ChatGPT, and add detail where missing. This would be an iterative process that could help a student consider holistic viewpoints towards developing an effective method for experimentation.

Sustainable product engineering and design

Subject Summary: This is a multidisciplinary subject that includes both engineers and scientists. It covers introductory level scientific and engineering design concepts including sustainable development, and product and processing line management in the chemical and food industry with an emphasis on sustainable manufacturing.

Assessment Types: The largest allocation of marks is with written research and reports activities totalling 50%. These activities are supported with presentations (5%), peer evaluations based on peer feedback and team contribution (15%), quizzes (15%) and participation (15%).

Result: For such a design-focused subject, and due to the suggestive nature of most of the assessment tasks, no final grade estimation is provided. Indicatively it would be very difficult for a student to pass the subject using ChatGPT due to the blend of assessment types addressing both technical and professional skills, including project-based learning tasks, team-based learning activities, reflective writing, and more.

Integrity Strength: Word length of the current version of ChatGPT was a major limitation and would make it difficult for students to use. To get around this, each section of a report had to be entered and scaffolded into ChatGPT separately. While ChatGPT provides general responses on design principles which are acceptable, the level of detail is lacking and thus, overall marks would reflect this. Students would need awareness of what a successful output would be and adjust the input accordingly in order to have a chance of passing.

Across most assessment components the biggest drawback was the generic response, lacking specific contextual knowledge. At times, it also made things up and provided inaccurate calculations. For improved results, ChatGPT requires substantial pre-training of information regarding the subject,

experimentation and workshop apparatus used. It is questionable if the time and effort would be time better spent than undertaking the work themselves.

ChatGPT struggled most in essay style questions, including reflective assessments where the students are asked to reflect on a particular class activity and the specific subject learning outcomes as well as program learning outcomes. With prompt engineering, the student could have ChatGPT reach a pass level for some of the activities.

An example of a prompt that would fail:

Reflect on the Sustainable Engineering Design subject – with a focus on industry site visits (in-person or virtual), guest lecturers and research into your own inspiring alumni

An example of a prompt that could move a student closer to a pass:

Reflect on the Sustainable Engineering Design subject – with a focus on industry site visits (we visited 'Masked' Brewery, saw process units and did site tour of plant), discuss your key learning from zero waste workshop in regards to sustainability and sustainable development goals, guest lecturers (by Professor 'Redacted' on P2X sustainable energy) and research into your own inspiring alumni Dr 'Redacted'.

As can be seen, the student would need a good understanding of how to engineer the prompt and have some idea of what the output should look like. However, the output was still rather generic and could not provide in-text referencing. A student would need to reflect on the weakness of the output and continue to adjust the input to get a much higher mark. Admittedly, it might be easier to build upon the ChatGPT output, than start writing from a blank page.

Integrity Weakness: From all the assessment tasks, the multiple-choice quiz component of the online quizzes was where ChatGPT excelled in terms of specific discipline knowledge topic areas (for example, the Sustainable Development Goals).

For the multiple-choice style assessments, any sufficiently general questions were answered correctly. With questions specifically related to subject content, ChatGPT's performance dropped to 50% correct. Originally, it was believed that using a tight time limit may constrain use. However, a browser add-on has been created that automatically solves all multiple-choice questions on a given webpage on the basis of ChatGPT's best guess for each question (Malko 2023). This plugin removes the benefit of imposing time constraints by removing the need for manual student input into ChatGPT.

Opportunity: While the output generated was not sufficient enough to pass most assessment tasks, it could help students with enough structural support that they could build upon to enable a passable submission. Using ChatGPT as a supportive tool, rather than an outright tool for cheating, would be a more appropriate use of student time.

Renewable energy and electrical power

Subject summary: This subject focuses on building students' knowledge and skills in renewable energy and is a final year subject for electrical power engineering students. A major project is undertaken, along with content-based assessment, and reflection on learning experiences.

Assessment Types: Assessment in this subject involves quizzes/tests (30%), lab work and reporting on the outcomes (10%), individual reflection on learning experiences (25%) with follow up interview (5%), and a comprehensive renewable energy design project involving a formal report (18%), progress presentation (4.5%) and final presentation (7.5%).

Result: Due to the assessment design of this subject, it would be very difficult for a student to use ChatGPT to obtain a passing grade with strong integrity observed in the lab, interview and design project tasks.

Integrity Strength: For the assessment activities such as lab report, interview, e-portfolio, report and presentation, ChatGPT could only provide support to a student and not a direct output that would lead to a passable answer. For the report, the word limit of the free version of ChatGPT required the assessment to be broken up into different components.

Overall, for many of the written components the output was very superficial. However, for a capable student the output provides the student with a framework that they could build upon to submit something that could pass. These tasks also rely heavily on software and simulations which cannot be directly accomplished within ChatGPT; however, guidance could be provided by ChatGPT to help students to use the software to achieve the outcomes of the assessments.

For the laboratory component, ChatGPT can't undertake the process on behalf of the student. Regarding the laboratory simulations, it can provide guidance on how to approach the simulation but can't carry out design and implementation in the laboratory software. For the submitted assessment task, it could only be used as an aid to help a student structure or write components of the report, primarily because the output it provided was too generic.

Integrity Weakness: ChatGPT's key strength was for the quiz questions, those containing simple calculations and explanations. In fact, the explanations for many of the questions were better than what would be typically provided by a student. However, the current version of ChatGPT struggled with quiz questions requiring translation from figures. This would result in missing components of an answer leading to partial marks, as everything else would be correct. An example of a question that ChatGPT can provide both a correct answer, together with very informative supportive description is shown:

Entered Question:

A residential PV system is to be designed to provide 10 kW of power and be grid connected. The proposed configuration will include a DC-DC converter and single-phase inverter. The inverter output is 230 V ac, 50 Hz, with input DC voltages possible in the range of 300–400 V. A boost converter is to be used in the implementation. The panel costs are given below: EN-300M-60-PERC1.pdf \$180 per panel SunPower300 - \$240 per panel KD150SX-UFU - \$90 per panel. For each provided datasheet: Determine the number of modules in a string, number of strings in an array, number of arrays, surface area for PV, weight, and cost. Which system configuration would you choose and why? Draw a one-line diagram of your system. What other factors may you need to consider in the design and implementation of a residential PV system?

Opportunity: While ChatGPT cannot complete the major assessment components for students, as identified above, it can provide guidance to a student in completing their assessments. The opportunity exists for students to utilise ChatGPT in order to learn how to develop their simulation model for example, utilise a new software tool, or write components of their reports with appropriate prompting.

Sustainable, environmental & social impacts of technology

Subject Summary: This subject focuses on developing engineering students' approaches to understanding the interactions between engineering and society from a philosophical, sociological and political perspective. Case studies of new technologies and engineering projects are examined to ground the learning in students' engineering contexts.

Assessment Types: Written assessment types (worksheets, summaries, feedback & recommendations) are the core focus on this subject (92%). Two visual activities (mappings) are used in support (8%).

Result: Based on an implementation that the ChatGPT output would be submitted without any editing by the student, ChatGPT would fail. If the student understood what was expected of them, they could modify the input to bring them close to a pass grade.

Integrity Strength: This subject is very research and writing-based and therefore very well targeted at ChatGPT functionality. However, in this subject ChatGPT struggled. The reason for this is that the subject requires a lot of integration with references, and external material that at times ChatGPT can't do, gets wrong or makes up. For example, even with a prompt to 'use at least one additional reference and give a citation in APA format' the paper cited had a title relevant to the topic but did not in fact exist, despite ChatGPT generating plausible sounding (but spurious) full reference details, including a (fake) doi address. Likewise, a prompt about summarising one section of the UNESCO

Engineering for Sustainable Development report evoked a response on 'section 3.7' which was unrelated to the actual section 3.7 from the source which is about 'Engineering and big data'.

Integrity weakness: Getting the context correct was a common problem and the output could be improved if the student worked on the input. For example, with one assessment the rubric requires specific reference to the learning resources from that week. However, if this was written out explicitly (i.e. 'refer to at least two of the following resources', and the links to the different resources were copied over) the response was quite strong. The references to the learning materials were a bit superficial, but definitely good enough to pass given the tight word limits on this task.

Many of the questions received a failure grade because the rubric clearly requested the integration of referencing which is, as was discussed earlier, something that ChatGPT currently is very limited at doing. However, this could probably change in future iterations if ChatGPT is trained with data from academic journals and conference papers or integrated with academic search tools like SCOPUS or Google Scholar.

However, if ChatGPT was used as a tool by a student, and the student had just enough knowledge to make appropriate modifications to the output, then ChatGPT could pass. For example, if the student took the foundational aspects of the output and manually inserted real references and corrected contextual mistakes, a pass mark could be awarded. In this case, the majority of the difficult work could be completed by ChatGPT, and less demanding editing undertaken by the student.

Opportunity: ChatGPT could be used as a tool to help introduce students to different ideas and perspectives. Critical thinking activities could be developed to analyse the given output to consider variables such as accuracy, bias and potential to shape public opinion. Artificial intelligence will have a big impact on society, and having students learn about its strengths and weaknesses is important.

Workplace practice & communication

Subject Summary: The subject has two components, one centred on career management and workplace skills, the other centred on teamwork solving multi-disciplinary real industry projects.

Assessment Types: Online quizzes reinforces theoretical concepts from the lectures, tutorials and subject reading comprised 5%, visual evidence of completed activities total 8%, extra-curricular activities in which students engaged in career building activities outside the course structure was allocated 10%, Oral activities such as presentations and tutorial discussions comprised 29%, with a range of written tasks (reflections, job-application, research, design reports) making up the remainder 48%.

Result: As a subject based with a strong communication focus, this was yet another candidate well suited for ChatGPT. There are clear assessment tasks that ChatGPT could not pass, those that it could, and those that it could provide enough guidance that students could build upon to pass with some effort. As a whole, it would be unlikely that a student could pass based solely on the use of ChatGPT.

Integrity Strength: The visual tasks (e.g. visual verification of a completed e-portfolio) had the best assessment integrity. The oral components (the fact that ChatGPT could not take the place of the student) and the design component could only receive partial help (outlined in opportunities below) supported assessment integrity.

Integrity Weakness: ChatGPT's strength was observed in simple activities such as quiz questions that were designed to support and remind students of theoretical concepts. However, quiz questions that required contextual understanding of the lecture or subject reading content resulted in many wrong answers. Beyond that, for many of the other assessment tasks like providing templates for all stages of the job application process, supporting students to write a reflection, and more difficult quiz questions, ChatGPT could produce some good output if the student worked a little on providing the correct input parameters. Providing the correct parameters would be the difference between a pass or fail. As long as the student understood what was

required, ChatGPT input could be manipulated to get the desired output. For example, compare the following:

Original Input:

Write me a job cover letter based on the following criteria: Adaptive skills (adapting skills and knowledge to various tasks) Foundation skills (oral, written, numeracy & problem-solving skills). Outline that I need to develop some skills.

Modified Input:

Write me a job cover letter. Outline that I have skills for the criteria Adaptive Skills (moved from India to Australia, high grades in both my undergraduate and postgraduate studies). State that I am working on areas that require further development in the criteria foundation skills (completing ENGG101 to improve my technical engineering communication skills and completing a masters in engineering management to improve my project management skills).

The original input would provide an output that was very generic, while the second provided an output that did a great job of integrating and building upon the evidence plugged in as parameters. The result would be something that would receive a passing grade.

There are a number of oral assessments in this subject. While ChatGPT cannot undertake a student presentation, it could help write a speech that a student could memorise to obtain a passing grade. If it was a Zoom based presentation the student could easily just read the presentation created by ChatGPT. Therefore, the impact on integrity was based if the assessment assessed delivery of content or student understanding.

Opportunity: When it came to design-based assessments, ChatGPT could not provide an output that would be sufficient. However, the student could gain substantial support if they broke the design into its varying components, and if they applied the right supporting parameters. While the project has a core design component, for educational purposes there are compulsory components on items such as ethics, safety, sustainability, human factors, etc. With good prompts for each factor, at times a good written response could be provided, or at the very least, enough information to direct the student in order to add additional research to supplement the information provided.

Let's consider an example prompt *For the project involving 'Company' and Precast Concrete Insulated Panels, what Australian standards should I consider?* ChatGPT would provide a very good list of Australian standards with good supportive explanations. However, the onus of the students would still be to consider if: the standard is real, the standard is relevant, and to delve deeper into what they need to know specifically about the standard in relation to their design problem. This activity itself, is a valuable learning experience and would work better than the current process used.

Therefore, for some components a pass is possible, dependent on the project and attributes. Most interestingly was its ability to talk and raise awareness of standards and risks that the great majority of students would have missed. Therefore, ChatGPT can be a supportive tool for design projects, but students need to review the output carefully, or they can incorporate contextual mistakes made by ChatGPT into their work.

Engineering research

Subject Summary: The focus is on the development of research skills. This includes supporting student to discover a research gap and scaffolding this to frame a research problem, developing a research design, design of data collection analysis and interpretation frameworks.

Assessment Types: Online Quizzes used to reinforce theoretical concepts taught in lectures and tutorials are worth 10%. An online presentation of the students work also has a 10% allocation. A visual submission of a mind-map representing the connection of ideas is worth 2%. The remaining 78% is attributed to a range of written assessment tasks scaffolded to guide students through the research writing process.

Result: Students could not use ChatGPT to pass this subject, but could use it as a supportive tool.

Integrity Strength: As was observed in the technology sustainability subject, ChatGPT currently fails when it comes to referencing material. It is clear that it has not been trained with a substantial database of scholarly material. While a student would fail in the subject due to its substantial scaffolding at each level, it is evident that in its current form it would fail any basic pedagogical approach to assessing a research project. The ChatGPT low word count is another temporary major limitation, making it difficult to engage and produce output of necessary size, requiring work to be broken into small chunks.

Integrity Weakness: The online quiz was the only identified weakness, but as the task was for reinforcement of knowledge, even if students used ChatGPT the same goal would be achieved. Questions based on general facts were generally answered correctly and often with excellent supplementary information. Questions that required context or understanding images were mostly wrong. Integrity would become a greater issue when ChatGPT's word count increases, it is trained with more scholarly content and gets referencing correct.

Opportunity: While the approach of asking ChatGPT to write the research report would fail, there are a number of ways students could use the technology to support them in research. For example, if students copy/pasted components of research articles and requested a summary, a rather good non-technical summary can be provided.

Some examples of basic prompts to aid learning would be:

- Can you rewrite the text I just submitted using English that a 10-year-old would understand?
- Can you summarise the research method so that a 10-year-old could understand?

With the right input parameters, ChatGPT could be a very supportive tool. To do this, students would need to train ChatGPT with substantial information. This included information on context and feeding in articles of value/relevance. Once this task is completed, ChatGPT could assist students in the generation of ideas, and structure of ideas. It could help students formulate possible research questions and also provide the student with some direction on how they could consider creating an experiment to answer the research questions. They could also write some paragraphs of text and then get ChatGPT to improve the quality of that text. Therefore, although ChatGPT can't do the research yet, it can support students to write and think about the research if asked and used strategically.

Some examples of basic prompts to aid learning would be:

- Can you rewrite the text I just submitted using academic English at the university level?
- Given the text I just submitted can you suggest a few possible research questions?

Assessment integrity

In the first component of this study, discussed in the previous section, the use of ChatGPT was benchmarked against individual subjects. The second component of the research study is to investigate the impact ChatGPT has on assessment implementations. In this section, an analysis is made of the impact on different assessment types by looking at each assessment across all subjects. It is important to highlight that the strengths and weaknesses are of value only at the time of writing, as ChatGPT continues to be trained and new releases are in the pipeline. More importantly, is the fact that the library of artificial intelligence tools continues to grow. With over 1300 different applications covering fifty different fields such as coding, copywriting, e-commerce, fashion and design available at the time of writing (Futurepedia 2023), the diversity of capable tools will only expand. In the coming years we should be prepared for the scenario that there will be a tool that could find weakness in any assessment task that is not invigilated correctly. As the technology adapts to users' needs, the academic community will also need to adapt its assessment approaches if it wants to ensure academic integrity.

It is also important to consider how likely it is that students may turn to an application like ChatGPT. If we consider the risk/reward relationship of cheating motivation (Noorbehbahani, Mohammadi, and Aminazadeh 2023) then if an assessment is able to be completed without invigilation we need to assume an integrity risk. This has been the reason why plagiarism tools like Turnitin have been heavily implemented at universities. For high-stakes assessments, like exams, universities have been turning to the aid of proctoring software that can provide a diverse range of monitoring solutions, but the more invasive, the greater the ethical questions become regarding implementation (Alin, Arendt, and Gurell 2023). However, there are generally work-arounds available to most monitoring solutions. For example, Dawson (2016) highlighted some old 'hacks' that could still work today, such as the use of virtual machines. Moreover, if a hack is not available today, there will surely be someone that will find a workaround. Therefore, even for invigilated online exams one can assume that some students may access ChatGPT if the risk/reward is a strong enough motivator. This competition between the increasing performance of AI tools and the software tools designed to thwart their misuse can be compared to an arms race, where the ever-strengthening power of ChatGPT and other chatbots must be matched by a comparable investment in the continual development of robust plagiarism-detection tools to ensure the ongoing integrity of assessment.

Online quizzes

This category includes all assessments tasks that are currently used in an online quiz format using an e-learning platform. Nine of the ten subjects examined in this paper used an online quiz as a form of assessment. This included a variety of question formats including multiple choice, true and false, matching items, calculation and text input. ChatGPT was able to pass all quizzes for all subjects, apart from one quiz which focussed on essay responses that required visual verification of module completions. Therefore, if we exclude that one subject implementation to consider more traditional quiz questions, it is clear that online quizzes are a risk to assessment integrity. Therefore, a rethink is needed on how they are used. This applies in particular for discipline-specific topic areas.

Result: ChatGPT has a clear advantage.

Integrity Weakness:

ChatGPT does not get all online quiz questions correct, but reliably enough to allow a student to pass. How well a student passed was determined by the focus of the question. Strong success rates were shown for questions that sought to test theoretical knowledge, and simple calculation questions (complex, multi-layered questions that allow for partial marks were not heavily used with such a quiz format). A common technique to enable integrity for calculation-based questions for online quizzes has been to randomise variables so that each student has a unique question and answer. The onus of such an approach being that in a time-restricted format, students could not share information as each question is unique. ChatGPT can easily answer such questions, at speed, and at a passing percentage. As the passing percentage will most probably only improve with further training, the value and integrity protection of using random variables is no longer fitting. One example of this rapidly changing environment is the newly developed browser add-on mentioned earlier that can use ChatGPT to answer all multiple-choice questions on a given webpage (Malko 2023). The integrity discussion is no longer about students getting answers from friends but getting answers from ChatGPT.

Integrity Strength:

In the short term, a number of assessment-design strategies have been identified to reduce the speed and success of trying to cheat using the current version of ChatGPT. The first option requires using questions that are based off figures and tables. Students need to translate the information contained within figures and tables to a format that ChatGPT can understand. This can be time consuming, and the student might not be sure if they are doing the translation correctly. If this is for a timed quiz, the complexity of translating the information for ChatGPT may be slower than answering the questions themselves. It should be noted that future releases may overcome such shortcomings, so this may be a solution with limited life. The second option requires creating questions with a specific

context, for example directly from lecture notes or a subject reading that may be out of alignment from a generic situation. Questions in which ChatGPT did not know the context, by not having access to the material, were generally wrong. Students getting such questions wrong can be a pathway to flagging student use of ChatGPT.

Opportunity:

Online quizzes were generally used as a low-stakes assessment, often repeated throughout the semester as a method to provide regular feedback to students on their progress. Sometimes they were used as a simple reinforcement of theoretical concepts, or to provide a guarantee that the students engage with some threshold concepts at least once. In such low assessment environments, the integrity risk of ChatGPT would be low, as the motivator is content engagement. There are other methods that can be applied to provide this content engagement/feedback. For example, online quizzes could be used on subject e-learning sites to unlock future content. For example, a threshold concept quiz at the end of each week, something that is ungraded, and that students can repeat an unlimited number of times, with a passing grade used to unlock the following week's content. With or without ChatGPT the student would gain feedback and content engagement.

The data suggests that the use of online quizzes beyond such low-stakes assessments is very risky. While short-term strategies to complicate the ChatGPT data input were provided above, a work-around may well be found in future releases. With solutions available to overcome online invigilation, there is substantial integrity risk in using online quizzes and such implementations should be substituted with alternative assessment.

Numerical (Assignments and exams)

This represents the collection of assessments tasks where the answers are numerical in nature (e.g. calculation based) and are completed in a written format (that is, not online).

In-person exams were tested as if they were available to be completed via ChatGPT. This was used by three subjects, over multiple implementations. The maths subject also had numerical based assignments. This category considers questions found in online quizzes as well as those that are typically more complex than those found in online quizzes, with multiple parts, and due to this added complexity may contain partial marks.

Result: ChatGPT is slightly more successful, and a pathway is set for further improvement.

Integrity Weakness:

Across six assessments for three subjects, ChatGPT passed four (one failed only marginally). This is still a strong performance from the AI software. Many of the passable marks came from questions that closely relate to those found in online quizzes. For multi-part, or otherwise more complex questions, ChatGPT gave some good explanations, but generally made an error within the working and gave the wrong answer. Due to the complexity of the answer, however, ChatGPT would have received partial marks. This was similar with the assignment questions and echoes a recent study by Frieder et al. (2023) who found that the Maths capability of ChatGPT struggled at the graduate level, being good at understanding the question but failing to provide correct solutions. Additionally, complex diagram-based numerical questions in the picture format cannot be used in the current version of ChatGPT. This is not surprising given the text-based focus of ChatGPT.

The maths assessments in this category were also tested with Wolfram Alpha. Both technologies had strengths and weaknesses, but if the best of both technologies were combined, the overall grade would be higher. Students could use ChatGPT for the working and Wolfram Alpha for the correct answer. With further training and development of the chatbot, it should be expected that a merger of the strengths from both platforms will become a reality.

Integrity Strength:

As with online quizzes, questions with references to diagrams or tables are troublesome for ChatGPT. The findings also show that ChatGPT is weak in some advanced mathematical areas used in engineering. In the short term, this provides a solution if the in-person tests were required to be undertaken in an online format. In the long term, it would be expected that the combined

strengths of ChatGPT and Wolfram Alpha would merge. Assignments in mathematics, therefore, pose a similar integrity risk as online quizzes, and should be discouraged as the main means to assess learning. To ensure integrity, the safest bet would be to continue using in-person exams, supported by multiple short assignments and online quizzes to encourage engagement with the content prior to the exam.

Opportunity:

Wolfram Alpha has already been seen by some as not a direct threat but rather a tool to aid learning (Necesar and Pospíšil 2012). That is, there is an opportunity to move beyond current punitive associations and develop new learning opportunities where AI software can be used to support student learning, rather than misrepresent it. If ChatGPT is used as a Socratic tutor (OpenAI 2023f), it could provide students with free and highly capable tutoring. This would open up tutor access to all, especially the disadvantaged that can't afford a tutor now.

Code submission

Only one subject required the submission of programming code.

Result: ChatGPT is a clear winner when it comes to coding assessments.

Integrity Weakness:

ChatGPT is generally very good at writing entry-level code and answering coding questions. At times the code may need some fine tuning, but the level of explanation of what the code does at every level can help students with this task. Therefore, for traditional introductory coding assessments, assessment integrity is at risk if not undertaken in-person. This is not straightforward, however, as students would presumably still have internet access and would require careful invigilation to ensure they weren't discreetly accessing ChatGPT.

Integrity Strength:

Beyond introductory programming, weaknesses were found with more complex activities. This is similar to the results found by Jalil et al. (2023) with ChatGPT being able to correctly answer unmodified student exercises, but not able to pass a software testing subject. The complex task unsuccessfully undertaken within this study required ChatGPT to open and analyse a CSV file. In the same way that translating figures and tables hindered success, adding requirements beyond ChatGPT's direct text-entry capability provides a short-term pathway to possible questions that can support assessment integrity if an online examination is required.

Opportunity:

ChatGPT provides an opportunity to rethink how coding is taught. Its ability to provide code with very effective instruction allows for a change in teaching pedagogy. With ChatGPT's exceptional proficiency in code generation, the automation of coding through this tool is anticipated to become more commonplace, with programming subjects prioritising the development of computational thinking skills. Furthermore, the insightful commentary provided by ChatGPT can be used as a tool to explain the basics of coding. It can also be used as a tool to help find and fix software bugs (Sobania et al. 2023), or to translate from one programming language to another (Carlsmith 2023). This could be used as a pathway for students to learn at a more complex level by recognising failure and learning to troubleshoot issues as they arise (Nikolic et al. 2022), which could support students to progress more quickly to more complex programming activities.

Oral

Oral based assessments comprising presentations, interviews, pitches and quality participation in discussion were used thirteen times across five subjects.

Result: Tied result, based on ChatGPT's ability to support oral assessment.

Integrity Weakness:

ChatGPT cannot take the place of students in oral assessments. However, depending on the activity, a possible pass may be an outcome if ChatGPT could provide the text that a student could memorise and repeat in a presentation. The greatest threat comes from external tools. For

example, Microsoft Copilot, a tool powered by ChatGPT, can generate slides and speaker notes for PowerPoint (Spataro 2023). At most, students would be required to tweak the presentation, and Microsoft Copilot provides all the tools to help the student undertake such tweaks.

Integrity strength:

Memorising text for a presentation might not be a negative outcome if the goal of the presentation was to give students practice and confidence in speaking. That is, if the rubric focussed on delivery over content, ChatGPT could be seen as a tool that can help students with structure and content. If the understanding of content was a priority, interviews retain strong integrity as an assessment tool due to the random probing from staff. While ChatGPT may help provide the student with possible questions that may be asked, it is unlikely that it could anticipate all potential questions from academic staff. However, if a student did try and prepare for such outcomes, it would be a supportive learning experience. Using participation marks to incentivise quality discussion in tutorials or workshops is another alternative identified in this study. As such conversations are unscripted and random, ChatGPT cannot be used to prepare responses.

Opportunity:

For some time, it has been recognised that greater support, practice and focus is needed to develop the oral communication skills of engineering students (Berjano, Sales-Nebot, and Lozano-Nieto 2013). The assessment integrity risks of other assessment types may encourage more oral assessment implementations (Rudolph, Tan, and Tan 2023). While presentations are common in engineering education assessment, the greater use of interview and marks associated with contributions to quality discussions in class can be of great benefit.

Visual

Five visual assessments were used across three subjects. There were two different types identified in the study. The first being the submission of visual documents such as a mind map or diagram. The other being visual evidence of the completion of some form of activity such as screen grabs of e-portfolio entries, or a link to a completed webpage.

Result: Visual assessments have a strong academic integrity advantage.

Integrity Weakness:

Only one assessment was provided with a pass, and this was primarily due to the assistance that ChatGPT gave to building the template diagram. In this regard, ChatGPT was used more as a supporting tool. As such, visual assessments are a positive, but it must be noted that there are a growing number of AI software tools that are capable of producing drawings.

Integrity strength:

The types of evidence discovered in this study were mainly seen as a strength for assessment integrity. Currently there is no AI app that can draw a screen grab of a portfolio, build and demonstrate a personalised working website, or design a research-based mind map. However, while such assessments may seem impervious against AI software, there are other means available that could enable student cheating. This could be manually doctoring or modifying existing images, portfolios or websites.

Opportunity:

While the visual skills of AI tools remain weak, visual assessment tasks tailored to particular contexts or topics remain an authentic assessment strategy.

Written (Experimentation-based)

Four subjects contained a written activity associated with experimentation or laboratory work. All experimental work was assessed via a written component.

Result: Tied outcome. ChatGPT cannot undertake laboratory work. However, if a student undertook the practical work and was able to collect data and gain some understanding of the experience, ChatGPT can provide students support in writing up their lab work.

Integrity Weakness:

ChatGPT cannot take the place of the student in experimental work. However, all subjects included a written assessment via a proposal or report. ChatGPT could pass some components of a report such as the introduction, method or reflection, but did need to be well guided by the input parameters to go beyond a response that was very generic. In doing so a good written response could be provided, or at the very least, enough information to direct the student towards a structure that they could adapt to make it passable.

Integrity Strength:

Experimental/laboratory work is a strength from an academic integrity perspective. If implemented correctly, it ensures students do the work, and understand the work. If ChatGPT is used as a supportive tool, it could enhance learning. While traditional hands-on work could be associated with strong academic integrity, simulated/remote options if designed to demonstrate 'doing' or knowledge gained by activity or experience, can also be favourable. However, while the focus is on AI integrity, there are other weaknesses associated with traditional delivery. This can include team experimentation in which a student passively tags along, or students copying and paraphrasing other students' work. This is an existing challenge with team experimentation assessments, independent of chatbot technology.

Opportunity:

The fact that ChatGPT can't replace the student in undertaking experimental work, reignites the importance of experimental work in engineering education. Experimental/laboratory work has been the backbone of engineering education since the beginning, but its importance has fluctuated over time (Feisel and Rosa 2005). The rise of ChatGPT is an important opportunity to reflect on the type and volume of experimental work conducted across the curriculum. While all the subjects in this study used a written assessment component, the contribution from ChatGPT was more applicable in a supportive role. The question is, why are so many assessments report focussed, especially given how Nikolic et al. (2022) found that academics see laboratory report writing as a low-ranking learning objective. There are over twelve different assessments types used within experimental and laboratory work (Nikolic et al. 2021), including non-written options such as presentation, demonstrations, observations and interviews that could neutralise ChatGPT and therefore retain integrity. If the type of assessments in the laboratory changed, and the weighting of assessments associated with experimentation was increased, this could then create a counterbalance of other assessment tasks with greater integrity risk from AI applications. Associating assessment from experimental work with a technical fail grade if the component is not passed, could be a valid safeguard of student learning.

Written (Project-based)

Four subjects contained a written assessment activity associated with project work. This included reports, engagement plans, scoping requirements, critical reviews and solutions. The success ChatGPT had was dependent on the requirements of the assessment task.

Result: Written, project-based assessments were more successful. ChatGPT, could provide help with components, but did not provide enough advantage overall.

Integrity Weakness:

ChatGPT could provide information representing a borderline pass if it was prompted with the right information for some of the report options. It can provide some good information on the topic area, but as mentioned previously citations and references are limiting factors. This is something that will probably be improved in future releases with better integration with academic sources. Even now, if the information provided by ChatGPT was manually improved upon by the student, strong outcomes could be produced.

Integrity Strength:

Knowing how to train ChatGPT through prompt engineering to obtain a passable answer can be difficult. The time and skills required to tailor the prompt could instead be used on completing the task itself, creating some imbalance in the risk/reward ratio of using ChatGPT. However, in the hands

of a ghost-writer trained in such a task (that is, contract-cheating), it could be valuable and lead to a strong illicit business model. At the moment, without suitable training, the responses are very generic and are especially limited if references are required. With a strong rubric and clear expectations, project-based assessment remains a strength.

Opportunity:

ChatGPT can help students with projects. It can help them with idea generation, information, structure, summaries and point them in the right direction for further investigation. For example, it can point them in the right direction to standards, ethical considerations, and workplace health and safety risks associated with their project. Project-based work could be a very suitable starting point to introduce students to how AI technology can be used as a tool to develop and flesh out solutions. As outlined by Lee et al. (2016), project-based learning has many valuable benefits, but providing direction and scaffolding is very difficult and impacts motivation and success. Furthermore, Valentine, Belski, and Hamilton (2022) found that computer-based approaches to idea generation do not impact students' self-efficacy compared to traditional approaches, and can be more appealing to students. We hypothesise that integrating AI technology into teaching approaches in this area could bring value.

Written (Reflective & critical thinking-based)

Five subjects tested ChatGPT against written assessment tasks that focused on reflective and critical thinking.

Result: ChatGPT was successful. For most assessments, a passable effort was possible with the correct input/training, and especially if the student applied just a little effort to build upon the generated response.

Integrity Weakness:

If a student knew what the output should look like, they could use ChatGPT to pass. Therefore, if the correct input/training is provided, ChatGPT could provide a passable output. Depending on the rubric it was important to be specific, such as 'reflect using the STAR framework'. Depending on the activity, ChatGPT could only give generic answers, but written in a format that a student with some understanding of expectations could modify and build upon.

Integrity strength:

With a basic prompt the output would be very generic. When scaffolding (e.g. a particular context or engagement with some external factors) was required, ChatGPT was not very useful. For other assessments, without the correct input/training provided to ChatGPT, most if not all assessments would have failed. This showcases that a student would need to have a good understanding of the expectations, in order to either modify the output or know what input is required.

Opportunity:

ChatGPT could be used as a scaffolding activity to help students think about the structure of reflective and critical thinking-based writing. For a student with no place to start, a beginning could be welcome. The danger being that in giving a starting point, ChatGPT arguably hinders the creativity and deep reflection process required for such writing.

Written (Research-based)

Three subjects tested ChatGPT against assessments focused on research-based writing. Synthesising the discussions so far, it is easy to predict that the outcome was that ChatGPT struggled to pass any of the assessment tasks. The key drawbacks are its inability to reference, its tendency to make up references, word limits with the free version, and its apparent lack of training with scholarly articles used in academia. However, we should expect that these limitations will close in future releases, if not with ChatGPT then certainly with some other AI package.

Result: ChatGPT failed when it came to research-based writing assessments.

Integrity Weakness:

If the research required in the assessment task was primarily a fact-finding mission, ChatGPT mostly did a great job. The limitation being references, but if the student managed to add some manually, arguably an easy task, strong marks could be achieved. For any written submissions, if students are required to provide some references published within the last 12 months, this may help with the integrity of the task as current version ChatGPT will struggle with references within this timeframe as the corpus it was trained on pre-dates 2022.

Integrity Strength:

At the time of writing, assessments that are research-based and supported with required referencing have high integrity. ChatGPT struggles to engage with scholarly articles and tends to falsify information. If components of research are scaffolded through various layers of activity (as found in the subjects examined in this research), the difficulty for ChatGPT to produce something valuable increases.

Opportunity:

Given that ChatGPT is a low risk for research-based writing, the authors identified several ways that the technology could be used as a tool. The suggestions provided are a low integrity risk but can substantially improve and help students scaffold their research skills. Some of the opportunities include:

Developing understanding: Especially for early-year undergraduates, research papers can be very complex and difficult for students to understand. Students can input sections of a research paper into ChatGPT and get a simplified summary that can aid their understanding.

Idea generation and structure: If the student can train ChatGPT with some of their initial thoughts, and research articles of relevance, then an output can be provided that supports the generation of ideas and structure that could be used for presenting their work.

Abstract: As mentioned above, ChatGPT is suitable for summary related work such as explaining complex ideas to a non-engineering audience. If the students work can fit within the word limits, it could also be capable of writing an abstract.

Research questions & method: If the correct training is provided, ChatGPT can present some suitable options or at least point students in the right direction for both research questions and a methodology to conduct an experiment to answer those research questions. The output is based on the input but provides a case of how teaching students to use ChatGPT can be of some value.

Writing Quality: For students that struggle with formulating ideas and putting them into words, ChatGPT can be very useful. Students can write paragraphs of text and then have ChatGPT provide an edit. In part, this is not much different from a student using a professional editor.

Recommendations for future research & practice:

In this paper, we have benchmarked the current performance of ChatGPT against a range of engineering education assessments. Although we have used assessments from a range of subjects, this coverage has not been exhaustive. There may be opportunities to fill in the gaps, as it were, by considering additional subjects covering the full extent of the engineering curriculum and different disciplinary specialisations. For example, most of the subjects considered in this study were at the undergraduate level, and so one potentially fruitful line of enquiry would be to evaluate postgraduate assessments in engineering from areas like management or professional practice. Certainly, given the nascent power of ChatGPT to respond effectively to assessment prompts, it would be remiss of engineering educators to not understand its potential and use the insights generated to improve the authenticity and validity of assessment practices.

We have also discussed various opportunities to incorporate ChatGPT into teaching practices and conjectured how this might support student learning. Given the novelty of ChatGPT, these remain mere conjectures, meaning that there is much scope to implement ChatGPT and evaluate its effect on student learning.

While the focus of this study is on a worst-case scenario in the event of improper use of ChatGPT by students, this may not actually be the case. We know some fraction of students find a way to

cheat, regardless of the assessment mode (Ellis et al. 2020; Alin, Arendt, and Gurell 2023; Noorbehbahani, Mohammadi, and Aminazadeh 2022), so we need to find ways to change behaviour. Through greater integration of ethics into the engineering curriculum, changes may be possible (Gwynne-Evans, Chetty, and Junaid 2021; Frigo et al. 2021). This opens an opportunity for research to tackle the ethical implications of integrating ChatGPT into the classroom.

ChatGPT-4 commentary

As outlined earlier, just days before this article was to be submitted, GTP-4 was released to the public. According to OpenAI (2023f), GPT-4 is claimed to substantially increase its performance, especially in the areas of maths and physics. Early studies backup such claims (Bubeck et al. 2023). As a result, limited retesting was conducted on the Physics exam. GPT-4 worked a little bit better on some of the questions, but on other questions, GPT-3 produced more accurate results. As a consequence, GPT-4 only saw a 2% overall exam improvement. Therefore, at this stage, we do not expect major differences across the other similar assessments. However, more concerning was the recent announcement that a Wolfram Alpha plugin is being made available for GPT-4 (Wolfram 2023). This is a game changer for the numerical components of engineering assessments. This will enable both correct explanations, together with correct calculations. This plugin warrants future research.

One of the weaknesses found for ChatGPT in the study was interpreting questions with tables, figures, or diagrams. Information provided by OpenAI (2023f) suggests this new version can accept a prompt combining text and images. An example is provided by OpenAI which shows that ChatGPT can recognise in an image that the wrong connector is plugged into a smartphone. We were not able to access GPT-4 with imaging to test, but this highlights the cautions recurring throughout the discussion that a solution to such a limitation would eventually be found.

The ability to remain up to date remains an issue with GPT-4. It generally lacks knowledge of events that have occurred after the vast majority of its data cuts off (September 2021). This still provides an opportunity for assessments to work within the bounds of more recent events.

A major limitation found in the study was one of word count. Our experience was that the limit was at about 500 words. The maximum length of the text that ChatGPT can generate depends on various factors, such as the specific implementation of the model being used, the available computational resources, and the complexity of the language used in the input (OpenAI 2023a). The low word count experienced, made working with some of the written assessment tasks tedious. According to OpenAI (2023g), GPT-3 had up to a 3,000-word limit, that has now been increased to 25,000 words. A larger word count will allow for easy importing and analysis of large articles and publications, overcoming some of the limitations found in the study.

For coding, it appears GPT-4 continues to excel. Examples have been found on Twitter that demonstrate using GPT-4 to create simple games from words and using basic sketches to build a website. This reinforces the commentary in the discussion in that it is not suitable to solely focus on keeping up in this arms race. We need to adapt and change the way we do things to get ahead. We also need to understand that using AI is going to be an important professional skill that will be integrated into daily work life (Spataro 2023). If we only try to just keep up, we will be wasting our time. We need to get ahead.

Conclusion

In this study, nine authors from seven Australian universities benchmarked their assessments against ChatGPT to develop an understanding of strengths and weaknesses across assessments used in engineering education. Analysis was undertaken both subject-by-subject and by assessment type. Strengths, weaknesses and opportunities presented by ChatGPT were discussed.

In a subject-by-subject analysis, ChatGPT passed three, failed five, and two subjects were too close to call. When it came to assessment types, ChatGPT passed four, failed three, and two types were

tied. Therefore, the discussion becomes more on the types of assessments used in a subject, rather than if ChatGPT is better at one subject area or another. From this, the following recommendations are presented:

- Online quizzes or exams with a weighting that favours a risk/reward ratio towards cheating should be stopped. It is clear that ChatGPT can pass most quizzes. Supporting web browser extensions can automate this process. Short-term workarounds have been identified in this study to make it difficult for students to use ChatGPT in such a format, but it is highly likely that updates will overcome such strategies.
- Short-term solutions include heavily using figures and tables that are hard for students to translate into textual input, asking questions with a particular context rather than generic expectations, interacting with external items such as evidence items like e-portfolios and drawings, and including external files that need analysis (such as a CSV file used in a coding exercise).
- Oral presentation assessments remain a strong strategy, in that ChatGPT can't participate on behalf of a student, and so oral tasks should play a larger role in assessment. However, ChatGPT can create scripted text that students can memorise, so the structure of the assessment should be flexible so there is some uncertainty about what direction the ensuing discussion and questions may take. Consideration is needed on the objective. If it is delivery, then presentation formats work. If it is understanding, then interview formats have higher integrity.
- The continued use of in-person exams with invigilation is recommended, especially when the risk/reward for cheating is high. However, it must be acknowledged that students still find ways to cheat with this approach, and it is not fool proof.
- ChatGPT cannot create visual elements such as mind maps (as per the context used in this study), CAD drawings and the like. However, we need to be mindful that other AI applications will build upon such features.
- At the time of writing, the mathematical capabilities of ChatGPT are limited (at best, only suitable for the maths-styled engineering questions given in online quizzes), but it has been identified that if its capability could be merged with that of tools like Wolfram Alpha, that would be a game changer. This is something recently announced (Wolfram 2023).
- Engineering experimentation/laboratory work is a safe integrity bet. Like oral presentations, AI technology cannot take the place of a student. This can encourage a shift to more experimental and laboratory learning. However, it is important not to be complacent, just because ChatGPT has limited reach, a passive student could remain hidden in team activities and plagiarise work. The integrity risk is currently in the type of assessment used in laboratory work, with reports being particularly weak. The diversity of laboratory assessment methods outlined in Nikolic et al. (2021) should be considered.
- ChatGPT struggled with project-based, reflective, critical thinking and research related writing-based assessments. However, with the correct training and question input, in many cases, passable output can be generated. Implicit in this is the notion that a student needs to have a good understanding of how to use ChatGPT and what the output should look like to be successful. Such a student would most probably be able to undertake the work directly. However, the risk becomes associated with Ghost Writers. Once the technique is mastered they can produce both quantity and quality, allowing prices to come down and shift the risk/reward balance for students.
- At the moment referencing is a major weakness for ChatGPT. However, GPT-1 released in 2018 was trained with 117 million parameters, GPT-2 released in 2019 with 1.5 billion parameters, while GPT-3 released in 2020 with 175 billion parameters. With this exponential growth in complexity, it is only time before this weakness disappears.

As engineers we need to embrace tools that make us work smarter and more efficiently. Just as we scaffolded the calculator as a tool in education, so we must with AI technology. This is simply a continuation of the monumental adaption of technology in education (Jackson et al. 2018). To hide it, we remove advantages to move forward competitively. We just need to understand how to do so ethically and safely. Companies like Microsoft are on the verge of taking ChatGPT AI technology mainstream via Bing (search engine) and Copilot (an AI integration with Office 360) (Spataro 2023). As all students now need to understand online collaboration through platforms like MS Teams, Zoom and Webex, students will soon need to know AI as well. Momentum is already afoot in the education space, as an example an AI tutor is being introduced to support learning through the Khan Academy (KhanLabs 2023). As a result, we have determined some initial opportunities to integrate ChatGPT into engineering education where academic integrity risk was lowest. Suggestions include:

- Using online quizzes with zero assessment marks to confirm threshold concepts and unlock future content.
- Consider flipped assessments and the like, where the activity is based on critical thinking activities that reflect and analyse ChatGPT output.
- Consider assessments that require creativity and go beyond writing. Such an example is illustrated in Nikolic and Raad (2021).
- Developing assessments where ChatGPT is used to undertake calculations, but the solution comes from the creativity of the student.
- Using ChatGPT as an education tool to teach introductory programming, and moving earlier to more complex tasks that go beyond easily developed solutions. Consider teaching how ChatGPT can be used to easily move from one programming language to the next.
- The greater use of interview-based assessment where the teaching staff probe students' understanding. This limits the ability of students to memorise responses.
- Encourage greater use of project-based and/or experimental/laboratory learning experiences where solutions are unique and require experience to document and explain.
- For projects and for research, ChatGPT can be an invaluable tool. While difficult to provide students a solution, it is powerful in providing them with editing support, direction, scaffolding and ideas that will give them momentum to complete the required activities. We need to learn how to embrace it.

This study has provided a baseline indication of ChatGPT performance on assessment that the academic engineering education community can benchmark with, and provided advice based on our experience on how to move forward. As a community we need to understand that the technology is improving rapidly, and we need to be prepared for a very different learning environment in the next twelve to twenty-four months.

Acknowledgement

ChatGPT was not a co-author of this paper. Information extracted or synthesised from ChatGPT is clearly referenced or highlighted as an example.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Notes on contributors

Sasha Nikolic received a B.E. degree in telecommunications and a PhD in engineering education from the University of Wollongong, Australia, in 2001 and 2017, respectively. He is a Senior Lecturer of Engineering Education at the University of Wollongong. His interest is developing career-ready graduates involving research in teaching laboratories, industry engagement, work-integrated learning, knowledge management, communication, and reflection. Dr Nikolic has been

recognised with many awards, including an Australian Award for University Teaching Citation in 2012 and 2019. He is a member of the executive committee of AAEE and an Associate Editor for AJEE and EJEE.

Scott Daniel is a Senior Lecturer in Humanitarian Engineering at the University of Technology Sydney, and serves as Deputy Editor at the Australasian Journal of Engineering Education and on the Editorial Boards of the European Journal of Engineering Education, the African Journal of Teacher Education and Development, and the Journal of Humanitarian Engineering. Scott uses qualitative methodologies to explore different facets of engineering education, particularly humanitarian engineering. He won the 2019 Australasian Association for Engineering Education Award for Research Design for his work with Andrea Mazzurco on the assessment of socio-technical thinking and co-design expertise in humanitarian engineering.

Dr. Rezwanul Haque is a Senior Lecturer specialising in Manufacturing Technology at the University of the Sunshine Coast. As an inaugural member of the AAEE Academy, he has contributed significantly to the academic community. In 2019, Dr. Haque served as an Academic Lead at the School of Science and Technology, overseeing the launch of two new Engineering programs and reviewing existing ones. His dedication to learning and teaching earned him the prestigious Senior Fellow status at the Higher Education Academy (UK) in the same year. His research focuses on Engineering Education and material characterisation through neutron diffraction.

Dr. Marina Belkina is Lecturer and First Year Experience Coordinator at Western Sydney University. She has taught various subjects and courses (Foundation, Diploma, first and second years of Bachelor's Degree, online Associate Degree). She has implemented numerous projects to support learning, including: Creating the YouTube channel Engineering by Steps, Leading the development of HD videos for the first-year engineering courses, Developing iBook for physics, creating 3D lectures and animations for Engineering Materials, and conducting research focused on exploring student's barriers to Higher Education.

Dr. Ghulam Mubashar Hassan is Senior Lecturer in Department of Computer Science and Software Engineering at The University of Western Australia (UWA). He received his PhD from UWA. He completed MS and BS from Oklahoma State University, USA and University of Engineering and Technology (UET) Peshawar, Pakistan, respectively. His research interests are multidisciplinary problems, including engineering education, artificial intelligence, machine learning and optimisation in different fields of engineering and education. He is the recipient of multiple teaching excellence awards and is awarded AAEE Engineering Education Research Design Award 2021.

Sarah Grundy is an education-focused lecturer at the School of Chemical Engineering, The University of New South Wales. Sarah predominantly teaches design subjects at all levels (undergraduate to postgraduate). Sarah has over 15 years of experience in Research & Development, Manufacturing, and project management in industry. Sarah's passion is to develop students to be credible engineers and make their impact in whatever industry through authentic learning practices.

Sarah Lyden completed her BSc-BE (Hons) at the University of Tasmania in 2011. From 2012 to 2015 she was a PhD candidate with the School of Engineering and ICT at the University of Tasmania. From March 2015 to February 2018 Sarah was employed as the API Lecturer in the field of power systems and renewable energy. Since 2018, Sarah has been employed as Lecturer in the School of Engineering. Sarah has been a member of the School of Engineering and ICT's STEM education and outreach team.

Peter Neal holds BE(Hons) and PhD in Chemical Engineering from UNSW. Peter Neal is an Education Focused lecturer and Academic Governance Coordinator with UNSW Chemical Engineering. He has been teaching undergraduate and postgraduate engineering students for more than 10 years, and has a key focus on developing his students' capacity in design, enquiry, and professional skills. As the first Education Focused academic in the School of Chemical Engineering and as a PVCSE Education Focused Champion, he works to develop the educational capacity of his colleagues in his School, the Faculty of Engineering and across UNSW.

Caz Sandison is an Associate Professor in mathematics and the Head of Students for Mathematics and Physics at the University of Wollongong. An award-winning lecturer, she has a keen interest in mathematics education, spending lots of time with students and working out more effective ways of teaching and assessing, while thinking about Topological Groups in her spare time. Caz is keenly interested in how to better recruit, train and retain maths and science teachers through collaboration between discipline experts (mathematicians and scientists) and educationalist experts.

ORCID

Sasha Nikolic  <http://orcid.org/0000-0002-3305-9493>

Scott Daniel  <http://orcid.org/0000-0002-7528-9713>

Rezwanul Haque  <http://orcid.org/0000-0002-8641-4479>

Ghulam M. Hassan  <http://orcid.org/0000-0002-6636-8807>

Sarah Lyden  <http://orcid.org/0000-0002-5364-6011>

Peter Neal  <http://orcid.org/0000-0002-8831-5327>

Caz Sandison  <http://orcid.org/0000-0002-5475-9640>

References

- Adamopoulou, E., and L. Moussiades. 2020. An Overview of Chatbot Technology. Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16, 2020. Springer, 373–383.
- Agomuoh, F. 2023. *The Best ChatGPT Alternatives (according to ChatGPT)* [Online]. DigitalTrends. Accessed February 19, 2023. <https://www.digitaltrends.com/computing/the-best-chatgpt-alternatives-according-to-chatgpt/>.
- Alin, P., A. Arendt, and S. Gurell. 2023. "Addressing Cheating in Virtual Proctored Examinations: Toward a Framework of Relevant Mitigation Strategie." *Assessment & Evaluation in Higher Education* 48 (3): 262–275. doi:10.1080/02602938.2022.2075317.
- Alua, M. A., N. K. Asiedu, and D. M. Bumbie-Chi. 2023. "Students' Perception on Plagiarism and Usage of Turnitin Anti-Plagiarism Software: The Role of the Library." *Journal of Library Administration* 63: 119–136. doi:10.1080/01930826.2022.2146445.
- Bani, B. S., and A. P. Singh. 2017. "College Enquiry Chatbot Using ALICE." *International Journal of New Technology and Research* 3 (1): 64–65.
- Berjano, E., L. Sales-Nebot, and A. Lozano-Nieto. 2013. "Improving Professionalism in the Engineering Curriculum Through a Novel use of Oral Presentations." *European Journal of Engineering Education* 38: 121–130. doi:10.1080/03043797.2012.745829.
- Bubeck, S., V. Chandrasekaran, R. Eldan, J. Gehrke, E. Horvitz, E. Kamar, P. Lee, Y. T. Lee, Y. Li, and S. Lundberg. 2023. Sparks of Artificial General Intelligence: Early Experiments with gpt-4. *arXiv preprint arXiv:2303.12712*.
- Carlsmith, D. 2023. *Generate MATLAB Code Using ChatGPT API* [Online]. Accessed March 2023. <https://au.mathworks.com/matlabcentral/fileexchange/125220-generate-matlab-code-using-chatgpt-api>.
- Choi, J. H., K. E. Hickman, A. Monahan, and D. Schwarcz. 2023. ChatGPT Goes to Law School. Available at SSRN.
- Dahiya, M. 2017. "Theoretical and Simulation Based Approach for Controlling Aircraft Longitudinal and Lateral Yaw Damper Movement Using PID Controller." *International Journal of Computer Sciences and Engineering* 5: 21–26. doi:10.26438/ijcse/v5i9.2126.
- Dawson, P. 2016. "Five Ways to Hack and Cheat with Bring Your own Device Electronic Examinations." *British Journal of Educational Technology* 47: 592–600. doi:10.1111/bjet.12246.
- Ellis, C., K. Van Haeringen, R. Harper, T. Bretag, I. Zucker, S. McBride, P. Rozenberg, P. Newton, and S. Saddiqui. 2020. "Does Authentic Assessment Assure Academic Integrity? Evidence from Contract Cheating Data." *Higher Education Research & Development* 39: 454–469. doi:10.1080/07294360.2019.1680956.
- Else, H. 2023. "Abstracts Written by ChatGPT Fool Scientists." *Nature (London)* 613: 423–423. doi:10.1038/d41586-023-00056-7.
- Feisel, L. D., and A. J. Rosa. 2005. "The Role of the Laboratory in Undergraduate Engineering Education." *Journal of Engineering Education* 94: 121–130. doi:10.1002/j.2168-9830.2005.tb00833.x.
- Frieder, S., L. Pinchetti, G. Ryan-Rhys, T. Salvatori, T. Lukasiewicz, P. C. Petersen, A. Chevalier, and J. Berner. 2023. *Mathematical Capabilities of ChatGPT*. Ithaca: Cornell University Library. arXiv.org.
- Frigo, G., F. Marthaler, A. Albers, S. Ott, and R. Hillerbrand. 2021. "Training Responsible Engineers. Phronesis and the Role of Virtues in Teaching Engineering Ethics." *Australasian Journal of Engineering Education* 26: 25–37. doi:10.1080/22054952.2021.1889086.
- Futurepedia. 2023. *FUTUREPEDIA: The Largest AI Tools Directory, updated daily* [Online]. Accessed March 9, 2023. <https://www.futurepedia.io/>.
- Gilson, A., C. Safranek, T. Huang, V. Socrates, L. Chi, R. A. Taylor, and D. Chartash. 2022. "How Does ChatGPT Perform on the Medical Licensing Exams? The Implications of Large Language Models for Medical Education and Knowledge Assessment." *medRxiv*, doi:10.1101/2022.12.23.22283901.
- GPTZero. 2023. *GPTZero* [Online]. Accessed February 16, 2023. <https://gptzero.me/>.
- Gwynne-Evans, A. J., M. Chetty, and S. Junaid. 2021. "Repositioning Ethics at the Heart of Engineering Graduate Attributes." *Australasian Journal of Engineering Education* 26: 7–24. doi:10.1080/22054952.2021.1913882.
- Hassan, O. A. B. 2011. "Learning Theories and Assessment Methodologies – an Engineering Educational Perspective." *European Journal of Engineering Education* 36: 327–339. doi:10.1080/03043797.2011.591486.
- Helms, M. M., and J. Nixon. 2010. "Exploring SWOT Analysis—Where are we now? A Review of Academic Research from the Last Decade." *Journal of Strategy and Management* 3: 215–251. doi:10.1108/17554251011064837.
- Hu, K. 2023. *ChatGPT Sets Record for Fastest-Growing User Base - Analyst Note* [Online]. Reuters. Accessed February 3. <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01/>.
- Hulman, A., O. L. Dollerup, J. F. Mortensen, M. Fenech, K. Norman, H. Stoevring, and T. K. Hansen. 2023. "ChatGPT-versus Human-Generated Answers to Frequently Asked Questions About Diabetes: A Turing Test-Inspired Survey among Employees of a Danish Diabetes Center." *medRxiv*, doi:10.1101/2023.02.13.23285745.
- Jackson, T., S. Nikolic, J. Shen, and G. Xia. 2018. Knowledge Sharing in Digital Learning Communities: A Comparative Review of Issues Between Education and Industry." 2018 *IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*. IEEE, 783–787.
- Jalil, S., S. Rafi, T. D. Latoza, K. Moran, and W. Lam. 2023. ChatGPT and Software Testing Education: Promises & Perils. *arXiv preprint arXiv:2302.03287*.

- Kanwal, S. 2018. "Engineering Students' Engagement with Resources in an Online Learning Environment." *INDRUM*. KhanLabs. 2023. *World-class AI for Education* [Online]. Accessed March 21. <https://www.khanacademy.org/khan-labs>.
- Khanna, A., B. Pandey, K. Vashishta, K. Kalia, B. Pradeepkumar, and T. Das. 2015. "A Study of Today's AI Through Chatbots and Rediscovery of Machine Intelligence." *International Journal of u-and e-Service, Science and Technology* 8: 277–284. doi:10.14257/ijunesst.2015.8.7.28.
- Lee, M. J. W., S. Nikolic, P. J. Vial, C. Ritz, W. Li, and T. Goldfinch. 2016. "Enhancing Project-Based Learning Through Student and Industry Engagement in a Video-Augmented 3-D Virtual Trade Fair." *IEEE Transactions on Education* 59: 290–298. doi:10.1109/TE.2016.2546230.
- Lei, L. 2023a. *AI Literacy for All - A Toolkits Series (ChatGPT - Public version)* [Online]. Accessed April 18. https://www.youtube.com/watch?v=Dq2_zisU6ww.
- Lei, L. 2023b. *How to Design Essay Assignments if Students can Access ChatGPT or Other Similar Tools* [Online]. Accessed February 18. <https://tl.hku.hk/2023/02/how-to-design-essay-assignments-if-students-can-access-chatgpt-or-other-similar-tools/>.
- Malko, V. 2023. *GPT4Moodle - Solve Moodle quiz with Chat GPT* [Online]. Accessed March 2. <https://chrome.google.com/webstore/detail/gpt4moodle-solve-moodle-q/lmaphpodbkdoaklifbnlofidiafkfdjm>.
- Merriam, J., and L. Kraige. 1993. *Engineering Mechanics: Volume 1, Statics*. Singapore: John Wiley & Sons, Incorporated.
- Murugesan, S., and A. K. Cherukuri. 2023. "The Rise of Generative Artificial Intelligence and Its Impact on Education: The Promises and Perils." *Computer* 56: 116–121. doi:10.1109/MC.2023.3253292.
- Necasal, P., and J. Pospíšil. 2012, October 24–26. "Experience with Teaching Mathematics for Engineers with the aid of Wolfram Alpha." *Proceedings of the World Congress on Engineering and Computer Science 2012 Vol I WCECS 2012*, 271–274. San Francisco, USA.
- Nikolic, S., and R. Raad. Developing Creative Digital Communication, Learning Resources and Evaluation Skills with YouTube and Self and Peer-Assessment. 2021 IEEE International Conference on Engineering, Technology & Education (TALE), 2021. IEEE, 473–478.
- Nikolic, S., M. Ros, K. Jovanovic, and Z. Stanisavljevic. 2021. "Remote, Simulation or Traditional Engineering Teaching Laboratory: A Systematic Literature Review of Assessment Implementations to Measure Student Achievement or Learning." *European Journal of Engineering Education* 46: 1141–1162. doi:10.1080/03043797.2021.1990864.
- Nikolic, S., T. F. Suesse, S. Grundy, R. Haque, S. Lyden, G. M. Hassan, S. Daniel, M. Belkina, and S. Lal. "A European vs Australasian Comparison of Engineering Laboratory Learning Objectives Rankings." In *SEFI 50th Annual Conference*, 2022. Barcelona, Spain: European Society for Engineering Education (SEFI).
- Noorbehbahani, F., A. Mohammadi, and M. Aminazadeh. 2022. "A Systematic Review of Research on Cheating in Online Exams from 2010 to 2021." *Education and Information Technologies* 27: 8413–8460. doi:10.1007/s10639-022-10927-7.
- Nov, O., N. Singh, and D. M. Mann. 2023. Putting ChatGPT's Medical Advice to the (Turing) Test. *medRxiv*, 2023.01.23.23284735.
- OpenAI. 2023a. (Apr 2023 Version) ChatGPT Prompt 'Does ChatGPT have a hidden limit of about 500 words or 4,000 characters?' [Large language model] [Online]. Accessed April 18, 2023. <https://chat.openai.com/>.
- OpenAI. 2023b. (Apr 2023 Version) ChatGPT Prompt 'Does ChatGPT use NLP?' [Large language model] [Online]. Accessed April 18, 2023. <https://chat.openai.com/>.
- OpenAI. 2023c. (Feb 2023 Version) ChatGPT Prompt 'How and why is GPT-3 better than GPT-1 OR GPT-2?' [Large language model] [Online]. Accessed February 19, 2023. <https://chat.openai.com/>.
- OpenAI. 2023d. (Feb 2023 Version) ChatGPT Prompt 'What Bias and Accuracy problems are associated with using ChatGPT have on engineering education?' [Large language model] [Online]. Accessed February 19, 2023. <https://chat.openai.com/>.
- OpenAI. 2023e. (Feb 2023 Version) ChatGPT Prompt 'What impact will ChatGPT have on engineering education?' [Large language model] [Online]. Accessed February 19, 2023. <https://chat.openai.com/>.
- OpenAI. 2023f. *GPT-4* [Online]. Accessed March 15. <https://openai.com/research/gpt-4>.
- OpenAI. 2023g. *GPT-4 is OpenAI's most advanced system* [Online]. Accessed March 15. <https://openai.com/product/gpt-4>.
- Popli, N. 2023. *The AI Job That Pays Up to \$335K—and You Don't Need a Computer Engineering Background* [Online]. *Time Magazine*. Accessed April 19. <https://time.com/6272103/ai-prompt-engineer-job/>.
- Rudolph, J., S. Tan, and S. Tan. 2023. "ChatGPT: Bullshit Spewer or the end of Traditional Assessments in Higher Education?" *Journal of Applied Learning and Teaching* 6: 1–22. <https://doi.org/10.37074/jalt.2023.6.1.9>.
- Sobania, D., M. Briesch, C. Hanna, and J. Petke. 2023. An Analysis of the Automatic Bug Fixing Performance of ChatGPT. *arXiv preprint arXiv:2301.08653*.
- Spataro, J. 2023. *Introducing Microsoft 365 Copilot – your copilot for work* [Online]. Accessed March 16. <https://blogs.microsoft.com/blog/2023/03/16/introducing-microsoft-365-copilot-your-copilot-for-work/>.
- Stokel-Walker, C. 2023. "ChatGPT listed as author on research papers: many scientists disapprove." *Nature* 18/1/23.
- Susnjak, T. 2022. ChatGPT: The End of Online Exam Integrity? *arXiv preprint arXiv:2212.09292*.
- Swiecki, Z., H. Khosravi, G. Chen, R. Martinez-Maldonado, J. M. Lodge, S. Milligan, N. Selwyn, and D. Gašević. 2022. "Assessment in the age of Artificial Intelligence." *Computers and Education: Artificial Intelligence* 3: 100075. doi:10.1016/j.caeai.2022.100075.

- Valentine, A., I. Belski, and M. Hamilton. 2022. "Influence of Using a pen-and-Paper or Computer-Based Approach on Engineering Students' Self-Efficacy During Idea Generation." *Australasian Journal of Engineering Education* 27: 16–25. doi:10.1080/22054952.2022.2030105.
- van Dis, E. A., J. Bollen, W. Zuidema, R. van Rooij, and C. L. Bockting. 2023. "ChatGPT: Five Priorities for Research." *Nature* 614: 224–226. doi:10.1038/d41586-023-00288-7.
- White, J., Q. Fu, S. Hays, M. Sandborn, C. Olea, H. Gilbert, A. Elnashar, J. Spencer-Smith, and D. C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering With Chatgpt. *arXiv preprint arXiv:2302.11382*.
- Wolfram, S. 2023. *ChatGPT Gets its "Wolfram Superpowers"!* [Online]. Accessed March 2023: <https://writings.stephenwolfram.com/2023/03/chatgpt-gets-its-wolfram-superpowers/>.

Appendix A

For all tables:

- Modification level required to pass:
 - o **N/A:** related to modification not being applicable (e.g. ChatGPT could not take the place of a student for an oral presentation)
 - o **Fail:** if a modification was applied, the output resulted in work that would fail.
 - o **None:** correlated to an output that would give a passing grade simply by copying the assessment task information/question directly into ChatGPT.
 - o **Minor:** related to some minor modification being needed to the input in order to get a pass grade. These generally resulted in the need to input specific parameters or instructions that provided better guidance of the structure and content required in the output. For example, simply rewording the question to use the prompt engineering structure instruction, context, input data and output indicator.
 - o **Major:** correlated to substantial modifications required to achieve a passing grade. In essence this would require a student that has a solid understanding of what the output needs to be and has the technical understanding to experiment with the input to achieve the required result.
- Wolfram Alpha Comparison (maths only):
 - o **0.** Results same as ChatGPT
 - o **1.** ChatGPT correct, Wolfram Alpha was wrong
 - o **2.** Wolfram Alpha correct, ChatGPT wrong.
- ChatGPT Pass/Fail:
 - o **Pass: At the** minimum a passing grade could be achieved (if assessment is not subjective, grade is provided).
 - o **Fail:** A passing grade was not achieved.
 - o **Component Pass:** While an overall fail grade was achieved, components of the assessment would receive a pass.
 - o **Possible Pass:** This would be a borderline pass/fail and conditional either on another assessment component or students touching up the answer (e.g. supplementing it with a real reference).

**Table A3.** First year physics.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Intra-Session Exam 1	25%	Numerical	Exam. In practice this is a closed-book exam conducted in class.	Pass	66%	Fail to Major based on question	Consists of five numerical questions on Uncertainty, Vectors, Linear and non-linear motion and Newton's laws. Full marks for most questions. Difficulty and wrong answers are associated with questions with diagrams. Providing a detailed description of the diagram in the prompt can give the correct answer, but the description needs to be well written, and the question required to complete the table should be also translated into the prompt text.
2	Intra-Session Exam 2	15%	Numerical	Exam. In practice this is a closed-book exam conducted in class.	Fail	36%	Fail to Major based on question	Consists of three data analysis questions. For one of the questions, the answer was correct without any formatting or modification of the questions. For the other questions, each requiring translation on a table into the prompt, only partially correct answers were obtainable.
3	End of session exam	50%	Online Quiz	Closed book exam conducted online.	Pass	70%	None	Consists of ten numerical questions on Units and Uncertainties, Linear Motion, Applications of Newton's Laws, Forces, Rotational Motion, Equilibrium and elasticity, Simple harmonic motion, Waves, Electric circuit, Fluid mechanics. As before most questions were correct or only partially wrong, with the greatest trouble coming from questions that required diagrams to be well described.
4	Practical	10%	Written	Laboratory report marked at the end of the class	Component Pass	N/A	Minor (Passing component)	Practicals are marked at the end of the class. If the students complete the practical analysis at home, they can pass the discussion question with the help of ChatGPT, but that component is worth only 10%. For the report the student needs to consider experimental data collection, completion of data tables, and calculations.

Table A4. First year maths.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	ChatGPT commentary	Wolfram Alpha comparison	Wolfram Alpha (WA) commentary
1a	Assignment 1	3%	Numerical	Full solutions to five questions on indices, inequalities, summations, straight lines and matrix multiplication.	Pass	70%	Minor	Syntax for the mathematical expressions, particularly matrices can be difficult. The actual calculations were often incorrect (written correctly, but not calculated correctly). Otherwise, the questions were entered as they appeared on the assignment.	2	WA can only answer calculation questions. I often had to just enter the calculation to get an answer. Answers given were always correct.
1b	Assignment 2	6%	Numerical	Full solutions to five questions on trig identities, solving polynomials with complex numbers, sketch a function, eigenvalues and eigenvectors of a matrix, inverse matrices.	Pass	50%	Minor	Perfect trig identity solution was given but it could not work in complex numbers. No sketch was provided for the function, but the discussion on how to get the graph was quite good. It calculated eigenvalues, and while the instructions for the eigenvectors were correct, the actual calculations were incorrect. In fact, clicking 'Regenerate Response' in the ChatGPT interface often gave a different answer.	2	WA could not simplify the trig identity, but calculated the answers to all other questions correctly. If a student knew WA was reliable, they could use ChatGPT to get the working and WA to get the answers, hence, full marks on this task.
1c	Assignment 3	11%	Numerical	Full solutions to four questions on implicit differentiation, parametric differentiation, cross product and dot product of vectors (and their interpretation), intersection of lines in 3 dimensions.	Pass	79%	Major	Minor adjustments to questions required, re-phrasing the crux of the question. Once again, the explanation of what to do was often very good. However, calculations were extremely poor. I even tried to correct it, but it continued to give incorrect answers (e.g. $4t/10t$ would simplify to $2/5t$) or $1.(1) + 3.(-3) + 2.(2) = -2$ when it should be -4 . If I didn't know the answers	1 or 2	This time, WA did not do so well as the questions were more complex. I had to know what I needed to calculate to get any benefit from using it. WA could not do parametric differentiation at all. WA also does not give any interpretation of results, whereas ChatGPT could interpret the result of a cross product or dot product of vectors.



2	Quizzes	10%	Online Quiz	Weekly quizzes worth 1% each consisting of four multiple-choice questions each quiz. Testing content from the entire session.	Pass	60%	Minor	came from AI, the working often got part marks for doing most things correctly. ChatGPT passed all 10 quizzes with marks of either 2 or 3 out of 4 for each one. For many of the incorrect answers, the result was simply wrong (not one of the distractors). There were one or two where a bit of common sense had to be used to pick the right answer from the ChatGPT solution, and a couple of times, the right answer came from asking ChatGPT to regenerate the solution.	1 or 2	For any calculation question, WA fared better than ChatGPT. For any analysis question, ChatGPT fared better than WA because WA does not interpret English (such as, 'Find the plane that contains ...') Given the nature of WA, the questions had to be modified a couple of times so that just the calculation was being requested.
3	Mid-Session Exam	35%	Numerical	Exam containing 8 True/False, 12 Multiple choice and 4 short-answer questions on Fundamentals, Complex Numbers, Functions, Matrices.	Fail	43%	Minor	In practice this is conducted face-to-face. True/False questions were very well done, 6/8 correct. Multiple choice questions were poorly done (3/12). ChatGPT often gave an answer that was not in the options. Short-answer questions where working was required gained marks for working only. Most answers were incorrect. It kindly worked out that $\sin \pi/8 = \sin \pi/8$ when asked to calculate the exact value of $\sin \pi/8$.	1 or 2	WA was not helpful for 6 of the 8 true/false questions. However, gave the correct answer for 9 out of 12 multiple choice questions. The other three were conceptual and WA is not good with those questions. While ChatGPT gave the correct working for the short answer questions, WA gave the actual correct answer.
4	Final Exam	35%	Numerical	Exam containing 6 True/False, 8 Multiple choice and 4 short-answer questions on Differentiation, Integration and Vectors	Pass	65%	Minor	In practice this is conducted face-to-face. This time, many of the answers to the true/false questions did not make sense. For one true/false question, it said it was false, but gave working to show it was true. Only scored 3 out of 6. Performed much better	1 or 2	WA performed worse this time, obtaining about 11/26. A combined effort using both tools would have ended with 20/26.

(Continued)

Table A4. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	ChatGPT commentary	Wolfram Alpha comparison	Wolfram Alpha (WA) commentary
								on the multiple choice with 5 of 8. For the short-answer questions this time, when there were algebraic questions, it seemed to perform better than numerical calculation questions. Those calculation errors persisted whenever there were numbers involved. The system could do implicit differentiation.		

Table A5. Introductory programming.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Weekly Labs	10%	Online Quiz	Covered Python fundamentals, writing functions, conditions and simple loop, strings and lists, and dictionaries	Pass	87%	1	There were 5 online quizzes (labs) contributing 2% each and containing multiple questions. Majority of questions require short programs to problems which are automatically graded online. All question were copy-pasted and ChatGPT performed well in most of the cases
2a	Project 1	15%	Code Submission	Required to write a program to read CSV file and provide statistical outputs	Fail	30%	Major (Components that it could complete)	Project description is substantially changed as original project sheet had tables and images. The project also required the analysis of a CSV file. The query was regenerated multiple times to get some output. The generated program had no syntax error which is something positive.
2b	Project 2	20%	Code submission	Required to write a program to read CSV file and provide statistical outputs but of higher complexity as compared to Project 1	Fail	0%	Fail	No solution was provided. Multiple modifications were made to the prompt, but all were unsuccessful. The requirements were too many and solution required to open and analyse CSV file which was not possible by ChatGPT.
3	Final Exam	55%	Written	Exam. In practice this is a closed book and notes exam.	Pass	100%	Minor	The final exam is of 2 h and there are 8 problems in the final exam. The marks allocation for problems range from 5 to 20 marks depending on the difficulty level. All problems of exam are checked. Difficulty target at an in-person closed book and notes exam. If undertaken by ChatGPT then full marks obtained.

Table A6. Manufacturing technology.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Theory based understanding	15%	Online Quiz	Calculation based questions	Pass	100%	Minor	Traditionally a technique of generating unique student questions based off the student number is used to ensure no two students are solving the exact same equation. With ChatGPT this uniqueness poses no major challenge to worded problems. The only questions requiring some prompt modification involved explaining a Table via text.
2	Assignment 1	20%	Written	Assignment questions requiring interpretation of figures and multi-level questioning	Fail	N/A	Fail	All questions required written translation of figures. This was not an easy task, and for all questions tested the correct output could not be achieved to a suitable level
3	Lab Report	15%	Written	Lab report based on students completing laboratory activities	Component Pass	N/A	Minor – Major (Passing Components)	ChatGPT is not capable of participating in the lab to see the metal cutting operation in the Lathe and Milling machine, nor capable of participating in the lab to measure the samples with appropriate equipment. However, if prompted with some context and information, ChatGPT can help create a foundation that can be used to write the report.
4	Final Exam	50%	Numerical	Exam. In practice closed book in-person	Pass	64%	Minor – Major	The examination is comprised of eight components, some of which have sub components. All the eight exam items were checked by ChatGPT. Success was based on question type. Many direct text-based numerical questions could be answered correctly. Difficulty arose when it came to translating tables and figures. Also, some questions required deeper conceptual understanding of manufacturing theory.

Table A7. Engineering laboratory.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Introductory Quiz	5%	Online Quiz	Consists mainly of MCQs with essay responses	Fail	15%	N/A – Minor Dependant on question	Simple copy and paste was passable for the MCQ. However, it could not complete the essay questions where it was asked to submit a screen capture showing it had completed training modules. Nor would it confirm that it had attended the lab induction. However, students could use ChatGPT to help answer questions while completing those modules.
2a	Experiment proposals	21%	Written	Experimental proposal	Component Pass	N/A	Major (Passing Components)	ChatGPT can't generate the required length of response (10 pages). Therefore, I got it to generate each section in turn. * Background – very generic, no literature citations, did not refer to the project context, nor explain the theory in mathematical terms, despite asking for it. * Planning – was ok, although lacked specific knowledge of the actual equipment and made some things up. Missed steps in the method. Lack of specific data, * Data Management and Analysis – generic response. * Project Risk Management – Probably passable, but not specific enough. * Project Management – Ok, but it took work to get it to correctly interpret the instructions for the Gantt chart.
2b	Proposal defence	9%	Oral	Interview about each proposed experiment	Possible Pass	N/A	N/A (Oral Presentation) Major (Provide script)	ChatGPT could probably generate a script answering potential questions. These answers would likely appear plausible, but it is unlikely that it could quickly provide sufficient detail in follow-up questions that students would appear to be sufficiently prepared to an experienced demonstrator.
3	Technical report	45%	Written	Technical report documenting the results of their experimental investigations and interpreting their meaning.	Component Pass	N/A	Major (Passable Component)	As with the proposal, ChatGPT is not able to produce a text of the expected length. Therefore, provided with prompts for each report section. * Introduction – too brief and did not engage with the experiment context.

(Continued)

Table A7. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
4	Final Seminar	20%	Oral	Students provide a team-based presentation reflecting on their learning over the course	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	<p>* Method – It generated a variety of responses to slightly adjusted prompts, but contained errors due to not having sufficient information on the experimental apparatus</p> <p>* Results – Insufficient. It did generate some data, but not the key information we expect and what it did report was wrong. Despite being prompted, it did not include statistical tests.</p> <p>* Discussion – This was ok, but not long enough, and did not cite sources when comparing against the literature. The statistics was not what we usually recommend.</p> <p>* Reflection – ChatGPT was able to successfully employ various reflection strategies. The content was reasonable but lacked concrete examples and depth of insight.</p> <p>I asked for a script, and suggested using Gibbs reflective cycle – which is not in the assessment guide. I also had to specify which experiments had been completed. The response included suggestions for what goes on each slide and was a surprisingly good start. As in other cases, it lacked detail and generally just restated the details in the prompt. However, it was able to allocate different aspects of chemical engineering theory to different experiments. Further, it made sure it addressed all the prompt questions by distributing them across Gibbs stages of reflection.</p>

Table A8. Sustainable product engineering and design.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1a	Class participation	5%	Written	Reflective writing on class activity/workshop	Pass	N/A	Minor – Major	With the correct parameters (that is, include workshop title, course title), response provided was appropriate (pass). But provided generic answers which were not specific to the workshop and/or course hence would not achieve more than a pass grading. As criteria of assessment stated to be beyond 50% the response needed to connect to course learning outcomes. ChatGPT only provided generic CLOs, not specific to this course, the student would need to include.
1b	Class participation	10%	Written	Reflective writing on guest lecturer and research on inspiring alumni	Possible Pass	N/A	Major	Did not pass as marking criteria requires student to provide in text citations and high quality references. Again, answers are generic. However, if student knew what an acceptable output looked like and could build upon output, possible to pass.
2	Summative quiz	15%	Online quiz	Summative quiz on preceding weeks	Yes	80%	Minor	Any questions that are general, the answers are 100% correct. Any questions that are specifically related to course content is 50:50%. Also note that if a time limit is used, depending on time allocated, could be hard to use ChatGPT
3a	Group project – Technical research task	10%	Written	Research and complete critical review of journal within 12 months	Fail	N/A	Fail	Provides overall generic answer, however it is not specific and more importantly does not include research of recent articles within 12 months (as per the criteria). No references to accepted literature.
3b	Group project – Preliminary report	20%	Written	Preliminary report on options of the project solution	Pass	N/A	Minor	Does provide good overall information on the topic area. However, the lack of high quality references would not enable a student to achieve higher marks.
3c	Group project – Pitch	5%	Oral	2-3minutes Pitch of the recommended option	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	While cannot present the oral interview on behalf of the student, based on the selection criteria responses in the previous component, it can provide a suitable transcript that the student can follow
3d		20%	Written	Final report of project		N/A	Minor	

(Continued)

Table A8. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
	Group project – Final report				Possible Pass			Can use ChatGPT response as initial starting point and expand on and present the results in the required format as per the criteria. However, references again would be the problem, limiting the mark to 50%.
4a	Team evaluations	10%	Written	Quantitative and Qualitative feedback to team members and their contribution.	Pass	N/A	Minor – Major	Again, while the feedback answers are generic, it provides students an excellent response in this case about the general principles of teamwork. If students used this, combined with their own reflections on specific team members, then they would be able to achieve an outstanding mark.
4b	Peer evaluation	5%	Written	Quantitative and Qualitative feedback to peer on their Pitch	Pass	N/A	Minor – Major	Again, not bad at providing generic feedback on both technical and presentation style. However, no specific connection to the course content and context but overall will pass easily if modified appropriately by the user. Also note: the task is the feedback is on 5 different individuals (so the student will need to still develop their own feedback that varies).

**Table A9.** Renewable energy and electrical power.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	In-semester tests	30%	Online Quiz	Online quiz on renewable energy fundamentals involving descriptive questions, calculations and design.	Pass	62.5%	Minor	Theory questions/descriptive questions mostly passed where these were general, where more specific detail was needed aligned with discussions in class it wouldn't have necessarily passed. For a photovoltaic system design question where the required information was set out clearly – i.e. calculate number of panels, size, weight etc. it performed very well. For Wind Energy Conversion system questions ChatGPT couldn't interpret images so initial calculation was wrong, but correct calculation process was exhibited for remainder of the question. Very good at definition-based questions.
2	Lab work and report	10%	Written	Students write up a final technical report covering the four lab sessions and their findings	Component Pass	N/A	Minor – Major (Passable Components)	Could write some sections of report for students with appropriate level of input from them, but not without completing the lab models and collecting the output data to discuss. For design-based lab tasks using simulation software, appears to be able to generate quite generic/limited design outputs but without the relevant simulations to support. Can provide guidance on how to approach the simulation in particular tools but without specific information the student would need to search for and input, it cannot do the design. Could be used to write sections of a report convincingly but students would need to have relevant inputs from completing the task.
3a	Individual interview and ePortfolio (portfolio part)	25%	Written	ePortfolio of unit experiences including what has been learnt, team experiences, project work contributions.	Possible Pass	N/A	N/A (ePortfolio) Major (Written component)	Asking ChatGPT to write 1500 words addressing the assessment criteria produced mixed results. While the specific elements like learning experiences in the unit were addressed very superficially (and without evidence) and would fail without understanding the output and modifying the input, the more descriptive components were able to pass on certain criteria. The use of the companion interview to this task could also help overcome superficial/generic

(Continued)

Table A9. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
3b	Individual interview and ePortfolio (interview part)	5%	Oral	Interview with questions taken from the student ePortfolio submission to verify it is their own work/their understanding of the subject.	Fail	N/A	N/A – Fail	<p>responses from ChatGPT on these criteria by probing student experience further to ascertain if they should actually pass on the criterion. Currently format for ePortfolio is very open, so further constraining this to ensure weekly reflection aligned with course content could help overcome some of the more superficial addressing of these criteria.</p> <p>Complementary task to the ePortfolio with a series of questions developed for each student individually to follow up on elements of their ePortfolio and enable them the opportunity to pass each criterion. While ChatGPT cannot sit the assessment for the student it could contribute some components towards their ePortfolio, however with extensive questioning in the interview student understanding could be probed to verify their own work. Extensive questioning including verifying that students can complete things they claim in their ePortfolio to a satisfactory standard required. A more rigorous interview than has been conducted in the past would be needed.</p>
4a	Renewable Energy Design Project (Report)	18%	Written	Scope requirements, complete design, report outcomes. Utilise software tools (System Advisor Model, HOMER Pro, MATLAB-Simulink)	Fail	N/A	N/A – Fail	<p>Putting in a modified version with greater task context led to a very superficial response on the design project with insufficient detail to demonstrate technical knowledge and application of appropriate tools. Without integration with the design tools wouldn't be able to develop a comprehensive solution. Can input a HOMER Pro design proposal output and ChatGPT will write report content based on this</p>

4b	Renewable Energy Design Project (Presentation)	7.5%	Oral	Presentation on design project outcomes	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	but doesn't meet the full requirements of the assessment task. Components of the report could be written at a superficial level and with false references. ChatGPT cannot participate in place of the student in project presentation but could provide talking points potentially aligned with their presentation and the project report.
4c	Renewable Energy Design Project (Individual progress Presentation)	4.5%	Oral	Presentation on the progress of design project	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	ChatGPT cannot participate in place of the student in project presentation but could provide talking points potentially aligned with their presentation and the project report.

Table A10. Sustainable, environmental & social impacts of technology.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Weekly Worksheets	30%	Written	Weekly worksheets covering topics such as engineering, society & technology; code of ethics; policy; stakeholders; sustainability;	Component Pass	N/A	None – Major	Difficulty and result was highly dependent on the specific weeks' activity and the requirements from the grading rubric. In many cases providing supportive parameters for context was needed, for others substantial training was required. Many problems were found with contextual errors, referencing problems (e.g. inventing references that don't exist) and aligning requirements with rubric. Superficial responses found at times together with many inaccuracies. However, at times, the response was suitable or could be used to support learning if the student understood what was wrong with the output and could manually fix.
2a	Summary of topic	6%	Written	Define a technology topic (e.g. Bitcoin, nuclear energy, artificial intelligence) and discuss some elements of its social complexity. [group]	Possible Pass	N/A	Minor	Very plausible response that identifies the key issues, however the cited references do not exist. Apart from that it's a very strong example. Student could pass if they manually added references
2b	Stakeholder mapping	3%	Visual	List and map key stakeholders in that technology. [group]	Fail	N/A	N/A	Very thorough list and description of key stakeholders. Chat GPT cannot produce the mapping on to the power-interest diagram (however this would be straightforward if the ChatGPT 'stakes' were used).
2c	Stakeholder consultation	6%	Written	Stakeholder engagement plan with those stakeholders in regard to that technology. [group]	Pass	N/A	Major	If the specific key stakeholders aren't mentioned, the response is thorough (in terms of consultation strategies) but not tailored at all to the stakeholders identified previously (and so a marginal pass). However, if the key stakeholders are listed and some text from the rubric is used to ensure the strategies are linked to the different stakeholders, the response is very strong.
2d	Stakeholder's stake	5%	Written	Explanation of each stakeholder's power and interest in the technology [individual]	Fail	N/A	Fail	Very strong text but citations of non-existent (but plausible sounding) references.

2e	Stakeholder persona	5%	Visual	Develop a persona using a template diagram featuring name & image, behaviours & actions, demographic details, and needs & pain points	Pass	N/A	Major	With some substantial modification of the prompt (i.e. adding text asking for dot points for the different quadrants in the persona diagram) the generated text is very strong. Would need to be put back into the template with some massaging to get full score.
3a	Incorporation of feedback	6%	Written	Table summarising how they have incorporated feedback into their final submission	Fail	N/A	Fail	The consistency required between the classroom experiences of feedback (some of which the teacher would have given or at least been present for) and evidence in the document means that tweaking the prompt is probably harder than actually changing the document itself.
3c	Topic summary	5%	Written	Summary of technology topic	Possible Pass	N/A	Minor	Fail only because of the requirement to give references, otherwise strong. Could pass if student manually added references
3c	Stakeholder stake	6%	Written	Tabulated summary of stakeholder's power and interest in the technology	Possible Pass	N/A	Minor	Fail only because of the requirement to give references, otherwise strong. Could pass if student manually added references
3d	Policy initiatives	18%	Written	Detailed explanation of several policy instruments that would serve the interests of their designated stakeholder with regard to the technology	Pass	N/A	Major	The length of text generated by any single prompt is limited. Therefore, since this section is quite long (~1000 words), the prompt could be adapted to generate a particular type of policy instrument, to serve the particular interests of a designated stakeholder with regard to that technology. If the prompt is 'hacked' in this way, the response is quite strong – except again for the references, which are plausible sounding but not all real. With a more specific prompt of 'propose a policy instrument using behavioural economics', the references were in fact real – but for some reason were not listed as citations in the text, just a bibliography at the end.
3e	Evaluation & recommendation	10%	Written	Tabulated evaluation of pros and cons of the proposed policy instruments and a recommendation of which one should be prioritised.	Pass	N/A	Major	With some hacking of the prompt to highlight rubric criteria and give the previously generated policy ideas, the evaluation is very thorough and gives a clear recommendation.

Table A11. Workplace practice & communication.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1a	Career Ready Skills	4%	Written	1. Reflective Letter (communicating skills still in need of development)	Pass	N/A	Minor	With the correct parameters (directly include examples of the skills associated with what I was good at, and what needed further development and how I was going to do that), response provided was appropriate (pass), including with regeneration
1b	Career Ready Skills	1%	Visual Evidence	2. Evidence of e-portfolio entries	Fail	N/A	N/A	ChatGPT cannot generate images of student entries of e-portfolio. However, this is something the student could add on their own
2	Moodle Quiz	5%	Online Quiz	Covers multiple choice, true/false, and matching questions predominately related to emotional intelligence topics from subject reading	Pass	58%	Minor	Struggled with questions that required context from the subject reading, but for questions without the need for context, did well. Could not analyse image-based questions, but if the image could be well described got answer correct.
3	Career Upskilling Evidence	10%	Extracurricular	Students have the option to participate in a different range of activities that provide certification	Not Applicable	N/A	N/A	Many different options, but for any activity that provided certification by online quiz, it may be possible to pass
4a	Job Application Process	1%	Written	Cover Letter	Pass	N/A	Minor	If the correct parameters are entered, a suitable cover letter can be created that hits the requirements of the position
4b	Job Application Process	1%	Written	Resume	Pass	N/A	Minor	A template is generated that provides a sufficient structure for the student to pass
4c	Job Application Process	4%	Written	Selection Criteria	Pass	N/A	Minor	ChatGPT did a great job implementing the STAR framework to provide a suitable response with basic parameters provided for evidence
4d	Job Application Process	4%	Oral	Oral Interview	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	While ChatGPT cannot present the oral interview on behalf of the student, based on the selection criteria responses in the previous component, it can provide a suitable transcript that the student can follow
5	Career Portfolio	3%	Written	Reflection with evidence	Component Pass	N/A	Minor (Passable Component)	With the correct parameters (interestingly relevant parameters used previously

								were added into the response) it produced a suitable reflection. However, it failed the entire task as it could not provide evidence of e-portfolio images, but this is something the student could add on their own.
5	Career Portfolio	7%	Visual Evidence	Students build a website	Fail	N/A	N/A	ChatGPT cannot design a website, but provides direction to the possible tools and resources to help build one
6	Class participation – First Session	5%	Oral	Frequency and quality of discussion in weekly tutorials	Fail	N/A	N/A	ChatGPT cannot participate in place of the student in tutorials
7a	Client Research	1%	Written	1. Reflection on how they feel on the industry project assigned to them	Pass	N/A	Minor	With only a few lines of text indicating how the student felt, a sufficient and highly rounded answer was given
7b	Client Research	4%	Written	2. Students needed to undertake preliminary background research on the industry client that they would be working with	Component Pass	N/A	Minor (Passable Component)	While not providing enough information to write the research report itself, the output does provide the student with substantial direction on what to do to find the information needed.
8	Presentation of the Industry project proposal	5%	Oral	Students in their teams present initial ideas of their project to the industry client	Fail	N/A	N/A	Creative and technical processes are required to come up with a solution. Project management tools such as Gannt Charts etc. cannot be created by ChatGPT. With the correct background information, for some parts of the presentation, ChatGPT may provide acceptable narrative, but not enough to pass as a whole
9	Reflection on Engineering Practices and Standards	5%	Oral	Students undertake a presentation on a random topic based on one of the guest lectures	Possible Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	While ChatGPT does not create the slides, nor can it present the oral presentation on behalf of the student, depending on the topic, it does provide a good transcript that the student can follow
10	Industry project – Final report	30%	Written	Student teams provide a very detailed report outlining the solution to the client problem.	Component Pass	N/A	Minor (Passable Components)	While the project has a core design component, for educational purposes there are compulsory components on items such as ethics, safety, sustainability, human factors etc. With good prompts for each factor, at times a good written response could be provided, or at the very least, enough information to direct the student in order to add additional research to

(Continued)

Table A11. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
11	Industry Project Final Presentation	5%	Oral	Students in their teams present initial ideas of their project to the industry client	Fail	N/A	N/A	supplement the information provided. Therefore, for some components a pass is possible, dependent on the project Creative and technical processes are required to come up with a solution. Project management tools such as Gannt Charts etc. cannot be created by ChatGPT. With the correct background information, for some parts of the presentation, ChatGPT may provide acceptable supporting narrative
12	Class Participation – Second Session	5%	Oral	Frequency and quality of discussion in weekly tutorials	Fail	N/A	N/A	ChatGPT cannot participate in place of the student in tutorials

Table A12. Engineering research.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
1	Progress Check	5%	Written	Literature Planner: Complete a template that guides students to extract understanding from the articles they read	Fail	N/A	Fail	Of the random journal papers selected, ChatGPT could only provide a summary based on article title and abstract. However, if the student can copy/paste text into ChatGPT they can generate some summaries that can help them complete some components of the template, but not enough to pass
2a	Introduction Chapter	0%	Written	Literature Planner: Completed template based on Assessment 1 (Needed for other components)	Fail	N/A	Fail	Fail, as per Assessment 1
2b	Introduction Chapter	2%	Visual	Mind Map: Draw a mind map connecting ideas between articles from the template in 2a	Fail	N/A	N/A	Failed for two reasons. 1. Failed scaffolding from previous activities. 2. Can't draw a mind map as per the context/requirements set
2c	Introduction Chapter	18%	Written	Introduction: Write a thesis introduction chapter based on 2a and 2b	Fail	N/A	Fail	Failed primarily due to scaffolding. Work-arounds were attempted, including reversing the scaffold, but no output was generated that met requirements. However, it can assist the student in the generation of ideas, and structure of ideas that can assist them in completing the activity
3a	Final Report	5%	Written	Abstract	Fail	N/A	Fail	The best result may have been possible if all the text could have been analysed by ChatGPT. Unfortunately, length limitations were a problem and text input had to be reduced/summarised. The output generated could provide a starting point but would require manual manipulation to pass.
3b	Final Report	3%	Written	Introduction	Fail	N/A	Fail	Fail, as per 2b
3c	Final Report	33%	Written	Literature Review	Fail	N/A	Fail	As per the introduction, a literature review as per requirements could not be satisfactorily generated. However, as per the introduction it could be used as a tool.
3d	Final Report	10%	Written	Research Design	Fail	N/A	Fail	While an output to expectations could not be provided, ChatGPT could lead the student to formulating possible research questions based on basic feed in of research information related to the topic. It could also provide the student

(Continued)

Table A12. Continued.

Number	Assessment name	Assessment weight	Assessment type (format)	Assessment description	ChatGPT Pass/Fail	Grade	Modification level required to pass	Commentary
3e	Final Report	0%	Written	Literature Planner (0%): Used as per Ass 1 and 2a	Fail	N/A	Fail	with direction on how they could consider creating an experiment to answer the research questions. Fail as per Assessments 1 and 2A
4	Presentation	10%	Oral	Presentation of research work	Component Pass	N/A	N/A (Oral Presentation) Minor (Provide script)	ChatGPT cannot present on the students' behalf. However, it can provide tips on how and what to present. It can also be used to generate various summaries that the student can use to simplify the content and structure the presentation
5	Critique	5%	Written	Students must critique research presentations from other students and compare against their own	Fail	N/A	Fail	While ChatGPT cannot critique the presentations, it can provide students with suitable structural advice on how to complete the activity
6	Online Quizzes	10%	Online Quiz	Questions related to theoretical and conceptual understanding of carrying out research	Pass	52%	None – Minor	Modification, depended on the question, but was mostly minimal. Questions based on general facts were generally correct and at times provided excellent supplementary information. Those that required context or understanding images were mostly wrong.