# ASCILITE 2024
## Navigating the Terrain:
*Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies*

# GenAI Teachers: Constructivist Learning Design and Value Propositions

**Chris Honig**
The University of Melbourne

Generative AI is increasingly used in higher education, prompting the need to effectively test, refine, and robustly evaluate its educational impact. Tools like Generative Pre-trained Transformers (GPTs) enable conversational interactions that support Constructivist learning designs: active learning strategies allowing students to engage deeply with course material through reflection and discourse. However, in our initial trials with an AI-tutors, students reported greater learning utility from non-Constructivist activities. This underscores the importance of not only developing new GPT tools but also educating students on their optimal use, designing assessments to foster appropriate learning strategies, and refining the tools themselves. We also explored self-directed learning scores but found no significant correlation with AI-tutor use-strategies, and only a slight preference for AI-tutors among highly self-directed learners. When using retrieval augmented generation (RAG) most learners could not distinguish between different large language models (LLMs) implying cheaper but refined models may be appropriate. Finally, we find AI-tutors offer a compelling value proposition to universities; student's perceive value of the AI-tutor exceeds the associated compute costs of running the AI-tutor. Similarly, students tend to prefer AI-tutors over similarly priced teaching alternatives.

*Keywords:* ChatGPT, GPT, GenAI, engineering education, AI-tutor, constructivist learning design, constructivism, self-directed learning

## Introduction

The use of chatbots in tertiary education pre-dates the release of ChatGPT-3 in late 2022 (Okonkwo and Ade-Ibijola 2021) but its release and the surrounding surprise in its capability has created a surge of interest in applications for generative AI (GenAI) in tertiary education(Adeshola and Adepoju 2023). A key advantage of generative pre-trained transformers (GPTs) are their ability to understand context-specific language and engage in conversation with high verisimilitude. This means GPT-powered chatbots now offer a tool to create meaningful dialogue to drive student learning and so many emergent GPT-teaching tools use this conversational interface. The key question is how to use GPTs to maximize student learning. GPTs may enable entirely new teaching methods (that were not possible before them) but for now, it is common to discuss them in comparison to existing teaching methods, as a replacement or alternative.

GPTs do not need to be accessed through a chat interface, but as this is the most familiar format, many initial teaching use-cases have focused on conversational formats such as AI-tutors and AI-roleplay (for example simulated patients) (Honig, Rios et al. 2023, García-Méndez, de Arriba-Pérez et al. 2024, Honig, Desu et al. 2024, Sardesai, Russo et al. 2024).

Among the AI-tutor use-case, a range of new tools have emerged: Khan Academy has recently released Khanmigo (an AI-tutor), OpenAI has partnered with Arizona State University to develop new GPT-enabled tutoring tools. Although these GPT-tools may ultimately develop into wholly new teaching use-cases, for now they are conceived within the reference framework of existing teaching methodologies, and so are widely discussed as teacher equivalents. What is also becoming increasingly clear is that the role of a teacher is diverse and includes many distinct, nuanced activities, so there are many different functions an AI-tutor can adopt, to replicate or extend upon, with varying degrees of success. For example an AI-tutor could focus on: specific information retrieval (like an advanced context-specific search function); study coaching (motivating students); pastoral care (providing support or simple guidance). An AI-tutor can generate quiz questions (create revision material), but it may also engage in a dynamic conversational quiz with students (a Socratic tutor) (Honig, Rios et al. 2023). Use-cases reliant upon consistent factual precision may encounter a number of implementation challenges due to 'hallucinations' (inaccurate or non-sensible information), but study coaching and Socratic tutor use-cases, that facilitate a student to engage in the learning (rather than delivering the factual learning) resolve many implementation challenges.

In this paper we seek to look at a range of different functions of AI-tutors and hope to better understand which use-cases offered students the greatest utility in their learning. We understanding this through Instructivist and Constructivist Learning Design (Social Constructivism) and also look at intrinsic student motivations to use the tools, using a self-assessment of self-directed learning.

## Theoretical Frameworks

**Constructivist Learning Design**

GPT-enabled chatbots offer promise to enable best-practice teaching methodologies into new classroom contexts that to date are constrained by practical limitations. One of these is the use of Constructivist Learning Design, that is often contrasted against Instructivist approaches to teaching. Instructivist learning design centres on the teacher, who serves as the locus of knowledge and guidance. In-person class time is usually designed for direct instruction (for example lectures, presentations and demonstrations). Information is transmitted to students uni-directionally, and then assessed in structured activities that set pre-determined knowledge goals (eg exams, tests, assessments). Feedback is provided by an instructor, emphasising mastery of specific pre-set knowledge outcomes(Hein 1991). These types of teaching formats are practical to implement at scale (for example if one lecturer is speaking to several hundred students) and also in online learning environments (for example in pre-recorded videos that can be viewed at any time).

By contrast, Constructivist learning design emphasises active engagement and student-centred learning. Within the paradigm of Constructivism, students do not learn effectively by passively receiving information and instead actively construct knowledge, from a base-level of understanding, by direct application of the ideas (often through direct discourse or personal reflection) (Hein 1991, Ertmer and Newby 2013, Narayan, Rodriguez et al. 2013). So within the Constructivism paradigm, learners are encouraged to actively construct knowledge with their peers and their learning environment. Learning activities often focus on collaborative projects, discussions and hands-on activities, that mirror real-world, authentic experiences. This allows for diverse interpretations and gives many pathways to knowledge acquisition.

The key differences in the two pedagogical approaches centre on (Narayan, Rodriguez et al. 2013):

1. Knowledge Acquisition: Instructivist methods focus on knowledge transmission, while Constructivist approaches centre on exploration and interaction with the knowledge
2. Teacher Role: In Instructivist methods, the teacher is central as the knowledge provider and evaluator, while in Constructivist method, the teacher functions as a facilitator to support students' independent exploration and inquiry.
3. Student Engagement: Instructivist approaches frame students as passive recipients of knowledge within structured activities to demonstrate mastery of pre-defined content. Constructivist approaches emphasize collaboration, discourse, reflection and the application of knowledge in authentic contexts.

Although many institutions may be slow to change, Constructivist Learning Design, and its complementary approaches like Active Learning, are widely considered best practice today (Felder and Brent 2009) but are often limited by resourcing or practical considerations.

GenAI offers the opportunity to better integrate Constructivist learning designs, at greater scale and in new classroom contexts (Cronjé 2024). For example in online learning and MOOCs (massive open online courses) course design is often reliant on Instructivist approaches, by virtue of the format. The teaching methods must be scalable creating a 'one-size-fits-all' learning design, while the student cohort may have individual and diverse needs (Crosslin 2016). For example, students may be studying at different times and areas, some are studying in blocks while others study part-time with competing commitments, online discussion boards don't allow for dynamic discourse, the time of the coordinator may be limited (in large classes) and online engagement is often low. These factors all make constructivist learning designs difficult to implement. But GPT-tools may offer pathways to resolve this, by creating an AI counterpart to allow students to actively construct knowledge: An AI-tutor in a Socratic teaching use-case or an AI-roleplay allowing students to directly apply their learning. In both of these cases, students can actively construct knowledge through discourse with an AI, that can be adaptive to their individual needs, and can still help steer and reinforce their understanding. This can also be delivered at low cost overheads (in ways that human tutors would not be able to do).

Central to this is understanding if (and how) students use AI-tutors for Constructivist learning approaches and what additional supports or learning environment design may be required to enhance this learning.

**2.2 Self-Directed Learners**

Where Constructivist and Instructivist learning designs focus on the learning environment created *for* the students, complementary research looks at learning created *by* students in their own study habits. The 'self-directed learner' is customarily contrasted against a 'teacher-directed learner'(Knowles 1975) and these definitions focuses on the learner: what, how and when the learner structures their learning (O'Shea 2003). A self-directed learner takes initiative and responsibility for their own learning process and actively seeks opportunities to deepen their understanding independently, demonstrating self-motivation in their educational journey. This is distinct from a 'teacher-directed learner' who learns reactively, by passively waiting to be taught. Note that self-directed learning is understood as a skill that can (and should) be developed among students and is viewed as a spectrum of practice (Knowles 1975). It is not that students are merely locked into a single category. It is also important to note that this is not only about learning outcomes: students with low developed self-directed learning skills more likely to experience frustration, anxiety and failure (Knowles 1975).

While all individuals have the capacity for self-directed learning, and also the capacity to further develop, the degree of individual development varies because of individual subjectivities and personal history (Williamson 2007). It is important for educators to be able to understand individual students' level of self-directed learning for its further development. For this reason the self-rating scale for self-directed learning (SRSSDL) instrument was developed, which is a 60-item validated survey for measuring the level of self-directed learning one possesses (Williamson 2007) based on the Delphi technique (Goodman 1987). In the SRSSDL instrument, respondents are asked to rate their response to a series of statements on a 5-point scale ranging from: Always, Often, Sometimes, Seldom, Never. Ratings are converted to numerical values, here ranging from 5-1 in order. A sample statement is: "I identify my own learning needs". Across the 60 individual questions, all responses are converted to numerical values and then summed for a grand total. Respondents are then divided into 3 categories based on score: 300-221 indicates high self-directed learning; 220-141 indicates moderate self-directed learning and 140-60 represents low self-directed learning (we did not have any respondents in the low category within the survey respondents).

**Research Hypotheses**

We want to understand productive ways for using GPT-based teaching tools. An AI-tutor may be used in many different and nuanced ways, some of which will be more effective, because of the learning design itself or because of the capability of the GPT.

Students are not a monolithic cohort; they are diversity in their backgrounds and preferred learning structures. So we also seek to understand how to effectively use GPT-teaching tools not only in aggregate but also at an individual level. Are there specific characteristics that indicate different preferences for GPT-teaching tools?

Finally, we want to understand the value-proposition to students: how useful are these tools really, when benchmarked to the costs of developing and running them? Universities will soon face real operational and resourcing questions about whether to create GPT-learning assistants for students and so need to understand the student-demand and learning utility in the tools, to benchmark to their costs.

With these broad goals in mind we set the following research hypotheses:

1. Preference for Constructivist interactions. The interactive nature of the AI-tutor allows for discourse, that is not traditionally available to students when studying (eg at home alone reading a textbook or doing tutorial problems). So we postulated that students would find it useful to engage in Constructivist learning activities with the AI-tutor, both for their utility in learning and also because the GPT has enabled study practices that were previously unavailable.
2. Stronger preference for Constructivist interactions among self-directed learners. GenAI offers many possible teaching formats for students. In our current use-case however, we expected more self-directed learners to have greater preference and engagement with Constructivist functions of the tool. By definition, self-directed learners have higher levels of intrinsic motivation and are more strongly motivated by learning (rather than extrinsic motivation like assessment) and so we expect they will gravitate to higher order learning goals within Bloom's taxonomy.

## Methodology

**Activity Design**

We built 4 different AI-tutors and made them available to students in a second-year undergraduate chemical engineering subject. Each AI-tutor was based on a different LLM: GPT3.5 turbo, GPT4, GPT4o and Claude. All 4 AI-tutors used retrieval augmented generation (RAG) with the subject notes (over 42,000 words of text). This meant the AI-tutors could answer specific questions related to the course, while a generic LLM like ChatGPT would not be able to. This chatbot format is sometimes pejoratively called a 'GPT-wrapper' meaning it uses a pre-existing LLM with a unique system prompt, but without finetuning. We have not built a custom model. This is simpler to build, but also means the AI-tutor format is equivalent to most University AI-tutor tools: it does use RAG but does not use fine tuning or a custom model. This makes the findings more broadly applicable.

The front-end user interface (UI) of the chatbot was built in Streamlit, with API calls made to the respective LLMs. This meant the custom tutors could be made available to students without them being required to create accounts with the LLM provider (for example they would currently need a paid OpenAI account to access GPTs, for access to custom chatbots). This was significant for our privacy impact assessment (the voluntary use of the tool and the absence of logging of any identifiable data meant it fell below the threshold for a PIA at our institution). It was also our preference for student privacy. We did not log any user information other than the number of unique IP addresses accessing the chatbot. We did not log conversation histories. The total number of API calls per day were logged in our GPT-usage dashboards, but these were not identifiable.

In class, time was reserved to demonstrate the AI-tutor and basic examples of use. Students were given general instructions on how the AI-tutor could be used and suggested activities, including: explaining concepts, interacting through quizzes, building revision materials, generating exam questions and facilitating study sessions. The examples included both Constructivist and Instructivist uses of the AI-tutor.

Students were then given the opportunity to use the chatbots over 2 weeks, in unsupervised activity, in the lead up to their final exam. Student feedback was collected via a survey, distributed after conclusion of the subject, to minimize perceived conflict of interest.

**Data Collection Design**

We first engaged students in discussion about the use of AI-tutors for the subject. These discussions and initial observations helped inform a list of possible ways to use the AI-tutor. These included both preconceived uses (how we thought it should be used) and also directly observed uses (how we saw students using it). We recorded 19 distinct use-cases for the AI-tutor. After developing this list, we mapped the individual use-cases to Constructivist (9 cases) or Instructivist learning design (10 cases). We developed these 19 distinct use-cases into simple Likert prompt statements, with the structure "It was useful for my learning, when using the AI-tutor to…" with the use-case then limited to a short phrase of 6 to 10 words. For consistency and generalizability to other work, the Likert responses were limited to: "Strongly Agree", "Agree", "Neither Agree nor Disagree", "Disagree" and "Strongly Disagree" but with an additional item "I did not use it this way". This last option (did not use) was an important control given the exploratory nature of the survey. Students are being presented with possible use-cases for the AI-tutor, but may have had no experience using it this way (because this use case did not occur to them or they did not believe it would be useful). Importantly, we want to limit the responses exclusively to students who did use it within this format.

We note that in this survey design, we are assessing student perceptions of the learning utility within each use-case (rather than directly measuring the actual learning utility, for example based on exam results). In this study, we use student perception of learning as a proxy for the actual learning. There are obvious ethical constraints around controlling for the use of the AI-tutor (for example only allowing half of the class to access the AI-tutor and comparing the final results). So we acknowledge this limitation of the study (we are measuring student perception of the learning utility).

We developed survey questions to understand the value proposition of AI-tutors to students and different preference for each of the AI-tutor LLM models. The AI-tutors were labelled FunCE Bot A-D, and it was not revealed which interface made API fetch requests to which LLM model. This allowed us to investigate if students could perceive a difference between LLMs (when it was not explicitly stated). We also asked about pricing directly, both how much students thought the tool would be worth, but also by benchmarking to other teaching activity that can be priced (such as extra tutorial classes, one-on-one tutoring or coordinator-led revision sessions).

Finally, we included the SRSSDL questionnaire, a standardized test for self-assessment of self-directed learning. The surveys were housed in Qualtrics.

We did not log student conversations with the AI-tutors. This was important within our study but for nuanced reasons. Firstly, this was done for student privacy and to create a 'low-stakes' atmosphere when using the tool. The tool is experimental and we wanted students to feel free to experiment with it. Surveillance of their conversations with the AI-tutor may make students reluctant to ask 'dumb questions' and can curtail and honest use of the tool. Secondly, within our research questions, there is no real need to collect student conversational histories. This is a subtle distinction, but we are not trying to understand how students used the AI-tutor. Instead, we are trying to understand how students used the AI-tutor that were beneficial to their learning. Observing how students used the AI-tutor is not methodologically valid: for example students could try for several days with unfruitful queries or interaction activities, only to have a lot of learning occur in a few simple questions later. So actual use of the AI-tutor can not be used as a proxy for productive use of the AI-tutor.

**Data Analysis**

The surveys were completed anonymously so we did not request respondent's details, but to prevent duplicate replies, Qualtrics does log IP addresses of respondents, which can be accessed. After closing the survey we first removed this data. We then converted Likert prompt statements to numerical values ranging from 1-5 (for strongly disagree through to strongly agree) as is standard practice (Johns 2010, Joshi, Kale et al. 2015). We note that it's common practice to present average Likert scores with standard deviations but this typically presents some confusion over interpretation: for example a Likert score of 4 is not twice as good as a score of 2 (it actually represents opposing sentiment) and the distributions are non-normal rather than linear.

Instead we prefer to look at percentage agree/disagree and so presenting the data this way helps to reinforce the non-linearity of the results.

For the SRSSDL we followed a standard data-handling approach, of first converting the responses to numerical values (1-5 for responses Never to Always) and added the totals to assess the final self-directed learning score of the respondent (Williamson 2007).

**Survey Validation**
Survey validation is often overlooked in SoTL research but is important for a range of reasons: it ensures reliability (the results are reproducible), internal consistency and accuracy (the items measure what they intend to measure) and test-retest reproducibility (respondents offer similar results if measured at different times). It also helps ensure validity across the theoretical constructs and content itself.

To validate our survey we began by reviewing similar studies. Likert instruments are widely used and so the processes for developing reliable prompt statements are well understood (Johns 2010, Joshi, Kale et al. 2015). The author re-read the statements at different times to check for ambiguity. We also tried using ChatGPT to review the statements for ambiguity and suggest phrasing to improve the clarity (these produced no major changes). Finally, we shared the prompt statements with other SoTL academics with varied backgrounds, to cross-check for any cultural-specific interpretations or ambiguities that may arise. The SRSSDL was previously validated and exists within the public domain (Williamson 2007). Finally we pilot tested the survey with a sample selection of students (not currently enrolled in the subject and so not within the final study group) to observe their responses and discuss their understanding of the statements. These processes led to several adaptions of the final survey but did not produce significant changes.

**Survey Distribution**
The survey was distributed to students within the subject via an LMS announcement, with a request for voluntary completion. In both the LMS advertisement and plain language statement, it was made clear to students that participation was voluntary and responses could be removed from the study later. Students were offered two $30 gift vouchers as inducement to complete the survey.

Of 83 students enrolled in the subject, 40 completed the survey but only 37 were complete (45% effective response rate). As the response rate represents almost half of all learners within the subject, this gives some confidence that the aggregated responses are representative of the complete cohort experience.

## Results

There was high reported use of the AI-tutor among students, which can also be confirmed against actual usage measured of the tools. Through Streamlit, we logged 103 unique IP addresses accessing the AI-tutor (5 of these are from academic staff). This left 98 unique IP addresses from a class of only 83 enrolments, but students can access through multiple devices, accounting for the usage above 100%. We think this confirms student reports of high uptake among the cohort.

From the release date of the AI-tutors until the date of the final exam (25 days total) we logged a total of 3230 API requests across all AI-tutors (meaning the total number of student queries). Among a class of 83 students, this equates to an average of 39 questions per enrolled student or just over 1.5 question per student per day, while the tool was available. All of these initial metadata logs are indicative of high student use.

**Instructivist vs Constructivist Learning Design Usage**
We have listed all the AI-tutor use-case activities included in our study (table 1 below). Each use-case activity has been mapped to Constructivist or Instructivist learning design. For example, activities that involve the AI-tutor summarizing or retrieving facts are associated with to Instructivist learning design, while activities that involve reflection, dialogue or conversation are mapped to constructivism. For each use-case activity, we have

recorded the percentage of respondents who reported they did not use the AI-tutor in this way. This may be because they did not think it would be useful or because it did not occur to them. Among the remaining respondents who did use the AI-tutor in this way, with have calculated the percentage who agree with the statement (agree or strongly agree within the Likert scale). This means that the 'didn't use' and 'agree' percentage responses can sum to more than 100%. Respondents who agree represents the percentage of respondents who did use the tool in this way and agree with the statement (it is not calculated from the total number of survey responses). We have then ranked the use-case activities from those with the highest agreement to lowest agreement.

Our hypothesis when initiating the study was that students would prefer to adopt Constructivist learning usage with the AI-tutors, given the conversational functionality of the chatbot. Instead of simply reading text or completing short questions, students now had the capacity to discuss concepts with the AI-tutor in-depth, following up and testing their learning. The results however, did not bear out this hypothesis. The top 4 most strongly preferred use-case activities were all associated with using the AI-tutor within Instructivist learning designs: Short definitions of key concepts, summarization, summaries and study guides and retrieving factual information. We do note that when directly comparing similar use-cases there was a preference for Constructivist activities: An AI-tutor can be used to generate a revision quiz or actively quiz a student on the material. We note that when framed as an active quiz (rather than generating a fixed quiz) students rated the activity more highly (5$^{th}$ most common preference). But this does not really change the outcome that many of the activities rated by students to be most useful for their learning, are aligned to Instructivist design.

Table 1

*A list of use-case activities mapped to Instructivist or Constructivist learning design*

| It was useful for my learning, when using the AI-tutor to… | Mapping | Didn't use (%) | Agree (%) |
|---|---|---|---|
| provide short definitions of key concepts. | Instructivist | 7.5 | 93.9 |
| summarize lecture notes or textbook content. | Instructivist | 25.0 | 92.3 |
| create study guides and module summaries. | Instructivist | 27.5 | 92.0 |
| provide factual information. | Instructivist | 5.0 | 91.2 |
| quiz me on the course material. | Constructivist | 37.5 | 90.5 |
| cross-check and critique my understanding. | Constructivist | 12.5 | 90.3 |
| act as a simple FAQ resource. | Instructivist | 17.5 | 89.7 |
| act as someone to talk to about the course material. | Constructivist | 20.0 | 82.1 |
| act as a study facilitator, leading a discussion. | Constructivist | 35.0 | 81.8 |
| create revision quizzes. | Instructivist | 52.5 | 80.0 |
| offer explanations for my incorrect answers on tutorial problems. | Constructivist | 32.5 | 78.3 |
| act as a personal study coach, to provide motivation and positivity. | Constructivist | 35.0 | 72.7 |
| provide general study tips. | Instructivist | 65.0 | 70.0 |
| prompt reflective discussions on my learning goals. | Constructivist | 52.5 | 66.7 |
| create practice exam questions. | Instructivist | 50.0 | 62.5 |
| facilitate group study sessions or discussions. | Constructivist | 70.0 | 62.5 |
| help set and track learning goals. | Constructivist | 65.0 | 60.0 |
| recommend additional readings or resources. | Instructivist | 65.0 | 40.0 |
| translate text into another language. | Instructivist | 75.0 | 16.7 |

**Self-Directed Learners**

At the end of our survey instrument, respondents completed the SRSSDL to assess their self-directed learning (Williamson 2007). Only 35 respondents completed this survey in its entirety (of 40 total responses). On the basis of these scores, students were divided into 2 classification categories: High (indicative of effective self-directed learning, 20 respondents) and Moderate (indicative of partial self-directed learning strategies, 15 respondents). Note that within the SRSSDL there is also a 3rd category for low self-directed learning but we had no respondents fall within this category.

We note that our population sizes were already small given the small number of students enrolled within the class and this again fragments the cohort. So we consider the results indicative but not sufficient to draw strong conclusions. Noting this limitation, we would like to understand learning utility and learner preference for AI-tutors. The results show a slight preference for AI-tutors among self-directed learners, but as the margin is small and the sample size is also small, we do not draw a final definitive conclusion. We instead hope to investigate this further in a future study.

Table 2

*Student aggregated preference when categorized into high and moderate self-directed learners*

| | **High Self-Directed Learners** | | | **Moderate Self-Directed Learners** | | | |
|---|---|---|---|---|---|---|---|
| Likert Statement | Agree (%) | Average Likert score | Standard Deviation | Agree (%) | Average Likert score | Standard Deviation | Difference in mean |
| I found the AI-tutor to be a useful learning tool. | 100 | 4.6 | 0.5 | 87 | 4.3 | 0.9 | 0.3 |
| I would like AI-tutors available for all my subjects. | 90 | 4.6 | 0.7 | 93 | 4.5 | 0.6 | 0.1 |

We also investigated preferences for use-case activities with the AI-tutor based on degree of self-directed learning. High self-directed learners expressed greater preferences for all use-case activities, but otherwise we could not identify a significant trend among the cohort. Both groups had strong preferences for the use-case activities we mapped to Instructivist learning design.

## Discussion and Further Work

**Effective use of GPTs**

This paper is part of an ongoing body of study to understand how to use GPTs most effectively in education. The capabilities of contemporary GPTs are without precedent and are evolving rapidly as newer models become available every few months. So there is an important research direction in iterating how these new tools may be deployed to optimize learning utility to students, in keeping with the principles of agile software development.

The most interesting results are often unexpected or surprising. We began this research project with a clear line of though: GPTs enable conversational interactions and so could create Constructivist learning opportunities, that were previously inaccessible to students studying at home alone. GPTs enable interactivity. It's also clear within the literature that this interactivity (for example through active learning design) enhance student learning (Hein 1991). But in deploying the AI-tutor, students reported the most valuable learning experiences were actually aligned to non-constructivist activities (for example simple information retrieval or generation of study lists). This is interesting: it's an unexpected result. The key question now is: why this has occurred? This is part of an ongoing study, but here we speculate about possible reasons for the observation, that can be generalized to broader research on GPT-enabled teaching.

- The tool itself. The AI-tutor itself can be designed for different levels of interactivity or passivity through the system prompts and through hard coding directly (our front-end UI Streamlit allows for specific text to be sent). We intend to experiment with different levels of activity from the AI-tutor in future iterations, to try to drive conversational learning activities to help more robust formation of the knowledge.

- Student training or direction. Just as GPTs in tertiary education are new for academics, they are new to students as well. People need to learn how to learn, and appropriate strategies for engagement with the content. We intend to embed learning instructions into the tool itself (the AI-tutor can describe how to use it effectively) but also discuss with students, appropriate learning strategies to try to use with GPT-enabled tools.
- Assessment driving learning behaviour. It is well understood that assessment designs and student learning are directly related(Hargreaves 1997, Gijbels, Donche et al. 2014, Carless 2015). The shape and form of the assessment changes students study habits, for example exam papers that demand fact recall incentivizes students to engage in rote learning. The AI-tutor was released about 3 weeks before the final exam. Anecdotally from student discussions, and what may logically follow, is that students predominantly used the tool in preparation for their exam. This summative assessment drives behaviour to maximize performance on the exam (rather than engage in more reflective learning practice). Students have been incentivized to practice past exam questions or draw on specific information relevant for the assessment. So we also intend to trial AI-tutors in different learning contexts, such as situations where there is no assessment (for example extra-curricular learning activities) or where the assessment is formative only. These alternative contexts may incentive user behaviour to engage with GPT-tools in more constructivist formats.

**Usage and Costing**

We created 4 different AI-tutors utilizing different LLMs: GPT-4 turbo, GPT-3.5 turbo, GPT4o and Claude-3 Opus. Student users were unaware of the underlying LLM of each AI-tutor. In the survey instrument, we asked for student preference. The most common response was ‘no preference’ but after this, the GPT-4 turbo model was most popular (confirmed from API calls). This indicates that many students are unable to perceive a significant difference in LLM when enhanced with RAG.

We also investigated price points for the AI-tutors and identified a tipping point at between $AUD 20-$AUD 50 per semester: below this price most students would use the AI-tutor, but above it most students would not. Tracking total token count across all 4 AI-tutors the costs were $AUD 1.98 for the 25 days of usage, or $AUD 7.7 across a full semester. This is below the value-based price set by students, meaning there is a compelling business case for AI-tutors within Universities (the utility value to students exceeds the current LLM compute costs).

## Conclusion

We built a custom AI-tutor using retrieval augmented generation for a second-year undergraduate chemical engineering course for understanding learning utility to students. In an unexpected finding, student self-reported the greatest learning utility in non-constructivist uses of the AI-tutor (for example defining key concepts, summarization, study guides and fact retrieval) as opposed to constructivist uses (for example discussing the key concepts from the course in detail with the chatbot). We also investigated students’ self-reported level of self-directed learning (through the SRSSDL instrument). We found a slight preference among self-directed learners for the open-ended use of the AI-tutor, but the measurable difference was minor. Finally, when costing out the use of the tool, we found it offers a compelling value proposition to students: cost-based pricing (compute costs associated with GPT-tokens) were less than the value-based pricing (price points identified by students) meaning the tool may be financially viable within Universities and may represent an appropriate allocation of teaching resources.

## References

Adeshola, I. and A. P. Adepoju (2023). "The opportunities and challenges of ChatGPT in education." Interactive

Learning Environments: 1-14. https://doi.org/10.1080/10494820.2023.2253858
Carless, D. (2015). "Exploring learning-oriented assessment processes." Higher Education 69: 963-976. https://doi.org/10.1007/s10734-014-9816-z
Cronjé, J. C. (2024). A Constructionist Approach to Learning with AI: An “Old” Solution to a “New” Problem? Future of Information and Communication Conference, Springer. https://doi.org/10.1007/978-3-031-53963-3_2
Crosslin, M. (2016). "From instructivism to connectivism: theoretical underpinnings of MOOCs." Current Issues in Emerging eLearning 3(1): 6.
Ertmer, P. A. and T. J. Newby (2013). "Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective." Performance improvement quarterly 26(2): 43-71. https://doi.org/10.1002/piq.21143
Felder, R. M. and R. Brent (2009). "Active learning: An introduction." ASQ higher education brief 2(4): 1-5.
García-Méndez, S., F. de Arriba-Pérez and M. d. C. Somoza-López (2024). "A review on the use of large language models as virtual tutors." Science & Education: 1-16. https://doi.org/10.1007/s11191-024-00530-2
Gijbels, D., V. Donche, J. T. Richardson and J. D. Vermunt (2014). "Learning patterns in higher education." Dimensions and research perspectives. London and New York: Routledge. https://doi.org/10.4324/9781315885438
Goodman, C. M. (1987). "The Delphi technique: a critique." Journal of advanced nursing 12(6): 729-734. https://doi.org/10.1111/j.1365-2648.1987.tb01376.x
Hargreaves, D. (1997). "Student learning and assessment are inextricably linked." European Journal of Engineering Education 22(4): 401-409. https://doi.org/10.1080/03043799708923471
Hein, G. E. (1991). "Constructivist learning theory." Institute for Inquiry. Available at: http://www. exploratorium. edu/ifi/resources/constructivistlearning. htmlS
Honig, C., A. Desu and J. Franklin (2024). "GenAI in the classroom: Customized GPT Roleplay for Process Safety Education." Education for Chemical Engineers. https://doi.org/10.1016/j.ece.2024.09.001
Honig, C. D., S. Rios and E. Oliveira (2023). "A Tool for Learning: Classroom Use-cases for Generative AI." The Chemical Engineer.
Johns, R. (2010). "Likert items and scales." Survey question bank: Methods fact sheet 1(1): 11.
Joshi, A., S. Kale, S. Chandel and D. K. Pal (2015). "Likert scale: Explored and explained." Current Journal of Applied Science and Technology: 396-403. https://doi.org/10.9734/BJAST/2015/14975
Knowles, M. S. (1975). "Self-directed learning: A guide for learners and teachers."
Narayan, R., C. Rodriguez, J. Araujo, A. Shaqlaih and G. Moss (2013). "Constructivism—Constructivist learning theory."
O'Shea, E. (2003). "Self-directed learning in nurse education: a review of the literature." Journal of advanced nursing 43(1): 62-70. https://doi.org/10.1046/j.1365-2648.2003.02673.x
Okonkwo, C. W. and A. Ade-Ibijola (2021). "Chatbots applications in education: A systematic review." Computers and Education: Artificial Intelligence 2: 100033. https://doi.org/10.1016/j.caeai.2021.100033
Sardesai, N., P. Russo, J. Martin and A. Sardesai (2024). "Utilizing generative conversational artificial intelligence to create simulated patient encounters: a pilot study for anaesthesia training." Postgraduate Medical Journal: qgad137. https://doi.org/10.1093/postmj/qgad137
Williamson, S. N. (2007). "Development of a self-rating scale of self-directed learning." Nurse researcher 14(2). https://doi.org/10.7748/nr2007.01.14.2.66.c6022

Honig, C. (2024). GenAI Teachers: Constructivist learning design and value propositions. In T. Cochrane, V. Narayan, E. Bone, C. Deneen, M. Saligari, K. Tregloan, & R. Vanderburg. (Eds.), *Navigating the Terrain: Emerging frontiers in learning spaces, pedagogies, and technologies*. Proceedings ASCILITE 2024. Melbourne (pp. 287-297. https://doi.org/10.14742/apubs.2024.1487

Note: All published papers are refereed, having undergone a double-blind peer-review process.

# ASCILITE 2024

## Navigating the Terrain:

*Emerging Frontiers in Learning Spaces, Pedagogies, and Technologies*