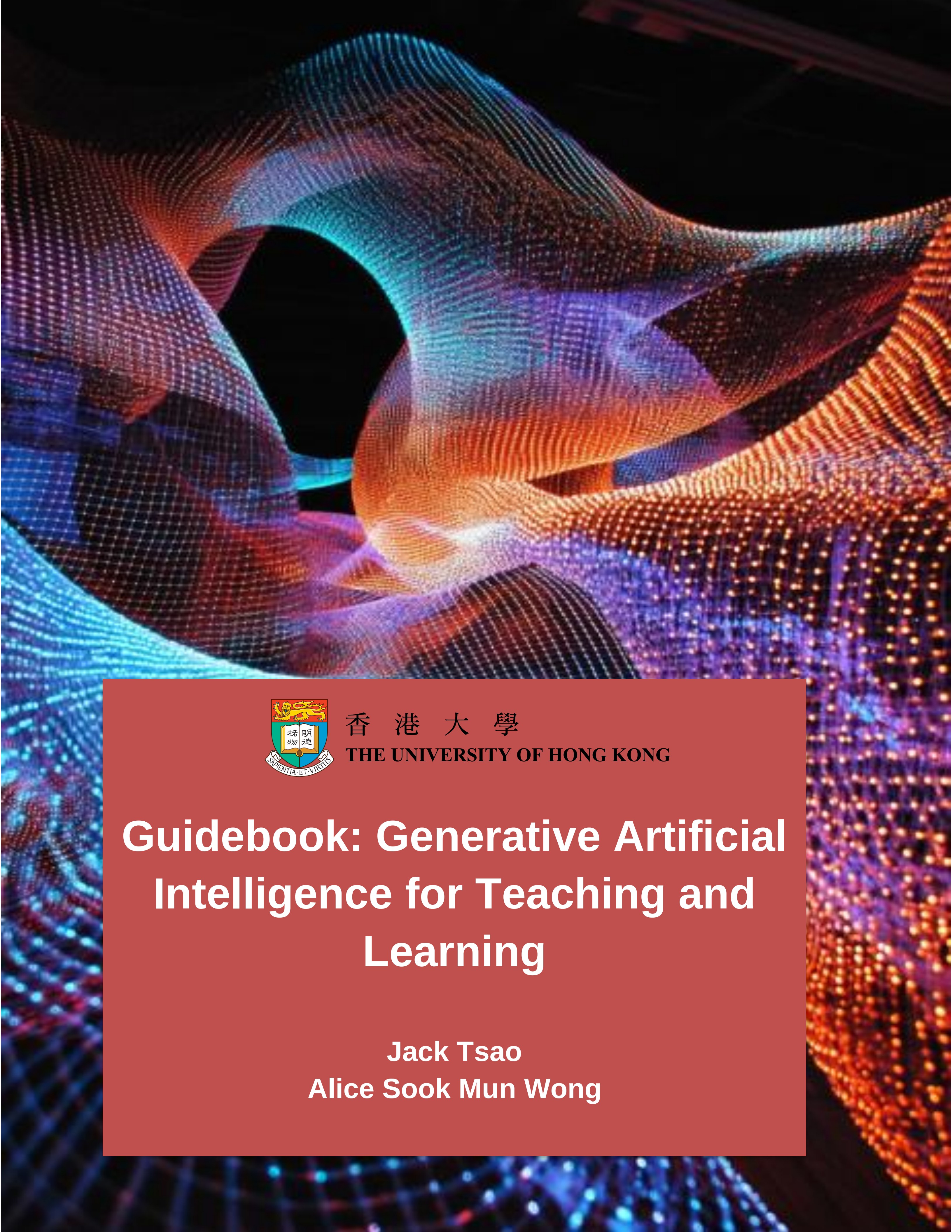


香 港 大 學
THE UNIVERSITY OF HONG KONG

Guidebook: Generative Artificial Intelligence for Teaching and Learning

Jack Tsao
Alice Sook Mun Wong

Table of Contents

Special Thanks **3**

GenAI in Higher Education *5*

Section I: Introduction to Generative AI **6**

1. Overview *7*

2. Understanding AI Models vs. AI Systems *8*

3. Categorising AI for Educational Applications *9*

3.1. Input and Output Modalities 9
3.2. Availability and Openness 10
3.3. Accessibility Considerations 11

4. Comparative Analysis of AI Models *13*

Section II: Perspective and Uses of Generative AI by Teachers and Students **14**

1. Overview *15*

2. Common themes from both teachers and students *17*

2.1. Verifying AI-generated output 17

3. Teaching and teacher perspectives *19*

3.1. How teachers use GenAI 19
3.2. Teacher Perceptions on Student Learning & Engagement 20
3.3. Teacher Perspectives on Limitations and Ethics 22

4. Learning and student perspectives *25*

4.1. Student Use of GenAI 25
4.2. Perceived Effectiveness and Benefits for Students 26
4.3. Student Perspective on Limitations, Risks, and Ethics 27

Section III: Application and Case Studies **29**

1. Overview *30*

2. AI- Embodiment and Role-Play in Humanities Ed *31*

3. Critical thinking through AI prompt engineering *33*

3.1. Critiquing AI responses 34
3.2. Structured Prompt Engineering 37

3.3. Interactive debating with LLMs 42

4. Case Study: Claude for Education 45

5. Case Study: Picturing China – AI vs. the World 46

6. Case Study: Future of Made-in-Hong Kong 46

7. Case study: GOAT - Who is the Greatest Economist of all Time and Why Does it Matter?" By Tyler Cowen 32

References 48

Authors 50

Special Thanks

Thank you to teachers, administrative staff, and research assistants who were involved in or supported this project (listed in alphabetical order of surname):

The University of Hong Kong

- Emily Chi Ting Chan, Common Core
- Anson Fan, Common Core
- Donn Gonda, Education
- Prof Gray Kochhar-Lindgren, Humanities
  Nicholas Mo, Centre for Applied English Studies
- Dr Michelle Raquel, Centre for Applied English Studies
- Dr Michael Rivera, Social Sciences
- Prof Julian Tanner, Common Core/ Medicine
- Penny Tse, Common Core

Other Institutions

- Dr Michal Camit, University of New South Wales, Australia
- Dr Danielle H. Heinrichs, Griffith University, Australia
- Prof Collier Nogues, Department of English, Chinese University of Hong Kong

Finally, we are grateful to the HKU Teaching and Development Grant (TDG), the Office of the Vice-President and Pro-Vice-Chancellor (Teaching and Learning), and the Curriculum Development and Quality Assurance Section (CDQA) of the Registry for their financial support of this project.

Figures

Figure 1. Student-Reported Benefits of Using GenAI 27
Figure 2. Sample prompt engineering structured exercise flow with 3 groups showing different levels of prompting. 40
Figure 3. Figure from Khan et al. (2024) illustrating a debate between two LLMs and a human judge. 43
Figure 4. From Anthropic's taxonomy for student-AI conversations, along with sample conversation topics based on those surfaced by Clio (Anthropic Education Report, 2025). 45

Tables

Table 1. Types of AI Models 13
Table 2. Table from Morris et al. (2024) depicting Levels of AGI. 15
Table 3. Teacher AI Use Cases 20
Table 4. Teacher Perceptions on Student Learning & Engagement 22
Table 5. Teacher Concerns: Limitations and Ethical Issues of GenAI in Higher Education 24
Table 6. Key Ways Students Use GenAI in Higher Education 25
Table 7. Summary of Student Concerns Regarding GenAI in Higher Education 28

GenAI in Higher Education

Generative Artificial Intelligence (GenAI) marks a significant turning point in the history of higher education by presenting unique opportunities and complex challenges that are reconfiguring relations between institutions, administrators, teachers, and students. GenAI systems that generate text, images, code, and other media from prompts are quickly being adopted by students and teachers in higher education, transforming the teaching and learning landscape.

As teachers, we should actively - but thoughtfully - explore, experiment, and engage with this technology to develop informed perspectives and strategies to harness its potential for student learning while mitigating its associated risks. These systems are already changing the learning experience, including offering instantaneous feedback, supporting writing and reading tasks, brainstorming and ideation, and study support.

For teachers, benefits include:

- Course design and planning
- Generating relevant and diverse teaching materials, including examples, case studies
- For assessment and feedback, including assessment and rubrics design, and grading and feedback
- Inquiry-based learning by assisting students in research tasks ranging from literature synthesis and data analysis to hypothesis generation

These can potentially free up teacher's time to focus on curating more impactful learning experiences and facilitate deeper student interaction and mentorship.

However, alongside these promising applications, universities and teachers need to devise policies and guidelines to navigate adverse outcomes of AI use. Issues such as fairness, equity, inclusion, lack of transparency, and overreliance can lead to unintended consequences that undermine the core mission of higher education in developing core competencies such as communication, collaboration, critical thinking, creativity, leadership, and intercultural understanding.

The guidebook explores some of these capabilities, presents examples of thoughtful applications, and engages in dialogue about GenAI's implementations and implications. By drawing on findings from the literature, alongside student and teacher interviews and case studies, this guidebook aims to provide teachers with the knowledge, frameworks, and practical strategies necessary to make informed decisions about GenAI use.

Our hope is for teachers to use AI not merely to enhance existing models of learning but compose with it as a means to unsettle current paradigms and assumptions in higher education teaching and learning to flourish in more creative and impactful trajectories.

Section I:
Introduction to Generative AI

1. Overview

This section will present the common vocabulary and baseline understanding of the various GenAI tools available commercially and open-source.

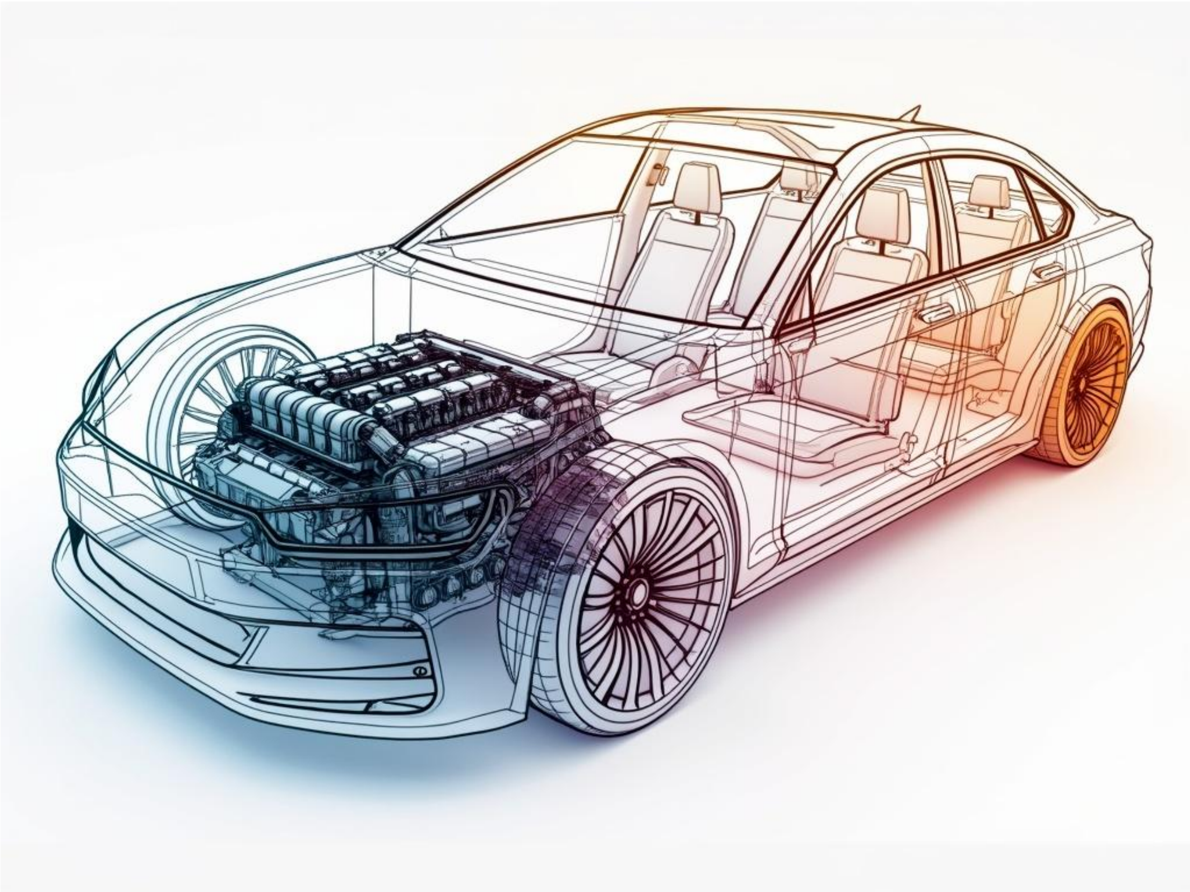
We explain the difference between AI models and AI systems, and categorise AI systems for educational use based on input/output modalities (e.g., text, images, audio, or multimodal), availability (open-source, open-weights, or commercial), and accessibility (cost, technical expertise required, hardware needs, and language support). These factors help educators select appropriate AI tools tailored to pedagogical objectives, technical capabilities, and resource constraints in diverse learning environments.

2. Understanding AI Models vs. AI Systems

AI Models are the underlying mathematical and computational structures trained on data to perform specific tasks. They represent the “brain” that has learned patterns from training data and can apply those patterns to new inputs.

AI Systems are complete implementations that incorporate AI models along with supporting infrastructure, interfaces, and tool functionality to solve real-world problems. An AI system typically includes one or more AI models, plus the surrounding user interfaces (e.g., a chat interface or coding virtual environment to execute code) that make the unified interface more pleasant to engage with by less technical users.

An AI model can be thought of as the engine in a car, while the AI and its surrounding support systems are the complete vehicle with the steering wheel, dashboard, seats, and other components that make it legible and steerable by most users.



3. Categorising AI for Educational Applications

AI systems can be categorised in several meaningful ways that impact their educational applications. The three most important factors to consider in education are how AI systems can vary in:

1. The types of information they can process;
2. The availability and openness of the system;
3. The accessibility of the AI system.

3.1. Input and Output Modalities

The most important distinction between AI models and systems are their input and output *modalities* - what types of information they can take in from the human user to work with (i.e., input type) and what sorts of content they can produce (i.e., output type).

- **Text-only**: Models that work purely with text are able to take in and generate textual conversations, essays, or documents. These are particularly valuable for essay feedback, content creation, and tutoring through conversational interfaces. Examples include early versions of OpenAI's ChatGPT models (e.g., HKU's own specialised HKU ChatGPT) and specialised writing assistants embedded within other applications (e.g. Copilot in Microsoft Word, Gemini in Google Documents etc.).

- **Image-processing**: Models that can process and generate images, making them valuable for quick visual prototyping or design projects. These enable creative assignments, visual analysis tasks, and can help create teaching materials. Examples include DALL-E, Midjourney, Stable Diffusion, Flux, Leonardo.ai etc.

- **Audio and Speech**: Models that work with audio and speech, enabling applications like pronunciation practice for language learning or music composition exercises. These can transcribe lectures, provide accessibility support, or help students practice language skills. Examples include Whisper or otter.ai for speech-to-text transcription, ElevenLabs for text-to-speech synthesis, and Suno or MusicLM for music generation.

- **Video**: Models that can analyse or generate video content. These can assist in creating educational animations, analysing recorded teaching sessions, or developing interactive video content. Examples include video generation tools like Runway ML, Google's Veo, or Hunyuan Video.

- **Multimodal**: Advanced models that combine these capabilities, processing multiple input types like images, audio, and text while generating outputs in one or more of the same modalities. These versatile systems can handle complex educational tasks requiring the integration of different content types. Examples include Claude 3, GPT-4V, and Google's Gemini.

- **Action-oriented (i.e., AI Agents)**: Experimental AI systems that can control computers directly by simulating mouse movements and keyboard strokes rather than relying on traditional application programming interfaces (APIs). These represent emerging tools that could potentially automate educational administrative tasks or provide assistance to students with disabilities. Early previews include OpenAI's Operator and Anthropic's Claude 3.7 sonnet (beta).

AI model's input capabilities do not always have to match its output abilities one-to-one. For example, while Claude 3.5 Sonnet can analyse images, it cannot create them. Similarly, some models can transcribe audio to text but cannot generate speech from text.

Model outputs can be chained (e.g., use text LLM to generate a better detailed prompt for image generator), creating powerful workflows for educational content creation. It can be hard to tell between chained models versus end-to-end multimodal systems. However, for practical educational purposes, it is not necessary to bring up this difference in the teaching and learning context unless the target of learning involves the additional chaining or unchaining of model outputs.

3.2. Availability and Openness

The second factor for educators to consider is the availability and openness of AI models and systems.

- **Fully open-source models** refer to models that allow researchers and technical users access to the model parameters but, more importantly, the (pre-)training data. Fully open-source models offer the potential for understanding how patterns (including biases) may have been transmitted from the training data to the model-generated output, which can be valuable for teaching critical AI literacy and allowing customisation for specific educational needs. These systems promote transparency but are unfortunately, rare, as many organisations choose to keep their datasets proprietary due to licensing, privacy, or competitive concerns. Examples of these include GPT-J from EleutherAI and DeepCoder from together.ai.

- **Models with open-weights** provide transparency about their parameters, making them modifiable with further training, but do not disclose the training data, making it hard to understand and probe potential biases or limitations that may exist within the model. These represent a middle ground between fully open and fully closed systems. Examples include Meta's Llamma models, some of the larger models from Mistral (e.g., Mistral 7B), and DeepSeek-V3 from DeepSeek.

- **Commercial AI models** are often closed and the most restricted, offering little visibility into the type of training data that went into creating and steering the model. However, commercial models are often coupled with delivery systems that are the most optimised for bulk delivery of output, speed, and ease of use for the non-technical user through clean web interfaces and apps. Examples include OpenAI's GPT-4 series, Anthropic's Claude series, and Google's Gemini series.

Models that are fully open-source or open-weights can theoretically be downloaded and run on a local device (e.g., one's own laptop or phone), unlike commercial, often cloud-based AI models, which are run on infrastructure owned by the respective commercial provider. The distinction between local AI systems and cloud-based ones has important practical implications for classroom use:

- **Local models** provide more confidentiality and privacy. Running these models does not require constant internet connectivity, which can be crucial for schools with limited bandwidth or strict data privacy requirements. This may be important for educators to consider when designing in-class assessments where the use of a computer is allowed, but the use of the internet is not. Local models can run on computers in school labs or even on student devices, enabling AI use without data leaving the premises. Examples include smaller versions of Llama, Mistral, DeepSeek, and other models optimised for local deployment.

- **Cloud-based systems** typically offer more powerful capabilities but require internet connectivity and may raise privacy considerations. These are currently the dominant form of AI systems in education. Examples include most commercial offerings like ChatGPT, Claude, and Gemini.

Some models can also be **fine-tuned on specific datasets**, allowing teachers to customise them for particular subjects or student needs - for example, training a system on grade-appropriate vocabulary and style (e.g. technical writing), or subject/course-specific content. This customisation potential makes AI more adaptable to different contexts and learning objectives.

3.3. Accessibility Considerations

Accessibility factors significantly impact which AI systems educators and students can reasonably use:

- **Cost and budget constraints** may affect which educators and students can access premium features. Many AI systems offer tiered pricing models with basic capabilities that are available for free but more advanced features that require payment.

- **Technical expertise requirements** vary widely across AI systems. Some require coding knowledge to implement, while others provide user-friendly interfaces accessible to non-technical users.

- **Hardware requirements** can limit deployment in resource-constrained educational settings. Some advanced AI models require significant computational resources to operate effectively.

- **Language support** varies between systems, with some offering robust multilingual capabilities while others primarily support English, creating equity issues in diverse linguistic settings.

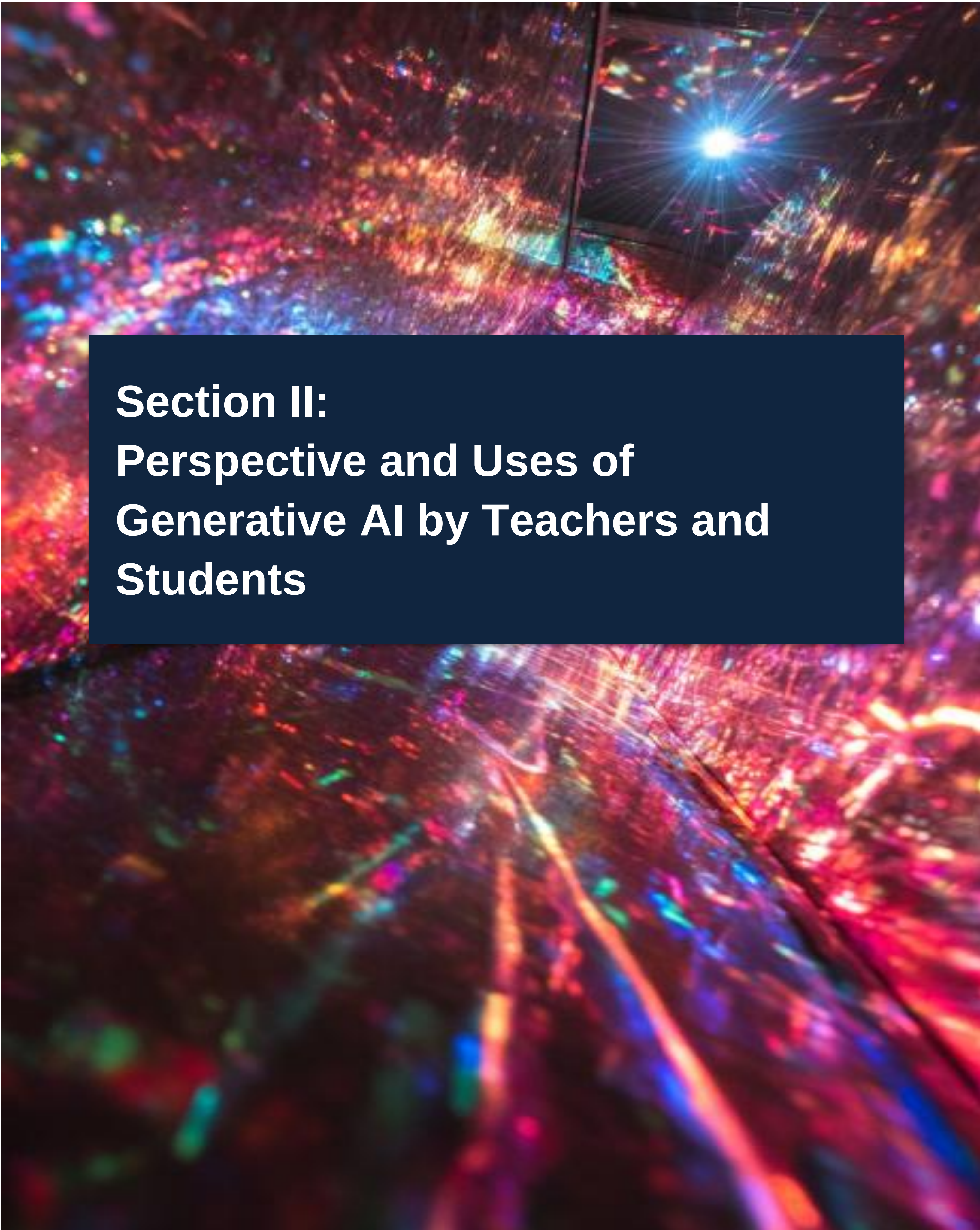
- **Disability accommodations** differ across platforms, with varying levels of screen reader compatibility, voice control options, and other accessibility features.

4. Comparative Analysis of AI Models

The table below provides a sample of modalities, applications, and features for teachers to identify which AI systems best suit pedagogical needs, technical capabilities, and institutional constraints.

Table 1. Types of AI Models

Model Type	Primary Input Modalities	Primary Output Modalities	Commercial Examples	Open-source or Open-weights Examples
Text-only LLMs	Text	Text	GPT-3.5, Claude 2	Llama 2, Mistral, Falcon
Image Generators	Text	Images	DALL-E 3, Midjourney, Adobe Firefly	Stable Diffusion, Kandinsky
Image Understanding	Images	Text	Claude 3, GPT-4V	LLaVA, CogVLM
Speech Recognition	Audio	Text	Whisper API, Google Speech-to-Text	Whisper, Wav2Vec
Text-to-Speech	Text	Audio	ElevenLabs, Amazon Polly	Bark, Coqui TTS
Multimodal Assistants	Text, images, documents	Text, sometimes code	Claude 3, GPT-4V, Gemini	LLaVA, GPT4All with vision capabilities
AI-assisted search engines	Text	Text	Perplexity, Gemini, ChatGPT with search, Elicit, Consensus	STORM (https://storm.genie.stanford.edu/)
Code Assistants	Text, code snippets	Code, text explanations	GitHub Copilot, Claude 3	CodeLlama, StarCoder
Video Analysis	Video	Text	Azure Video Indexer	TorchVision models
Educational Tutors	Text, sometimes voice	Text personalised feedback	Khan Academy Khanmigo, Duolingo Max, Claude for Education	Customised open LLMs for education



Section II: Perspective and Uses of Generative AI by Teachers and Students

1. **Overview**

Although current AI tools are not yet considered to be Level 2 (“Competent”) artificial general intelligence (AGI) in the sense of generally matching or exceedingly at least the 50th percentile of skilled adults (Morris et al., 2024), the Teaching and Learning space in higher education occupies a unique niche as universities represent one major institution where its (very human) students straddle the blurry line between being a “skilled” and “unskilled” adult.

Table 2. Table from Morris et al. (2024) depicting Levels of AGI.

Table 1. A leveled, matrixed approach toward classifying systems on the path to AGI based on depth (performance) and breadth (generality) of capabilities. The assignment of example systems to cells is approximate. Unambiguous classification of AI systems will require a standardized benchmark of tasks, as we discuss in Section 5. Note that general systems that broadly perform at a level *N* may be able to perform a narrow subset of tasks at higher levels. The “Competent AGI” level, which has not been achieved by any public systems at the time of writing, best corresponds to many prior conceptions of AGI, and may precipitate rapid societal change once achieved.

Performance (rows) x **Generality** (columns)	**Narrow** *clearly scoped task or set of tasks*	**General** *wide range of non-physical tasks, including metacognitive tasks like learning new skills*
Level 0: No AI	**Narrow Non-AI** calculator software; compiler	**General Non-AI** human-in-the-loop computing, e.g., Amazon Mechanical Turk
Level 1: Emerging *equal to or somewhat better than an unskilled human*	**Emerging Narrow AI** GOFAI (Boden, 2014); simple rule-based systems, e.g., SHRDLU (Winograd, 1971)	**Emerging AGI** ChatGPT (OpenAI, 2023), Bard (Anil et al., 2023), Llama 2 (Touvron et al., 2023), Gemini (Pichai & Hassabis, 2023)
Level 2: Competent *at least 50th percentile of skilled adults*	**Competent Narrow AI** toxicity detectors such as Jigsaw (Das et al., 2022); Smart Speakers such as Siri (Apple), Alexa (Amazon), or Google Assistant (Google); VQA systems such as PaLI (Chen et al., 2023); Watson (IBM); SOTA LLMs for a subset of tasks (e.g., short essay writing, simple coding)	**Competent AGI** not yet achieved
Level 3: Expert *at least 90th percentile of skilled adults*	**Expert Narrow AI** spelling & grammar checkers such as Grammarly (Grammarly, 2023); generative image models such as Imagen (Saharia et al., 2022) or Dall-E 2 (Ramesh et al., 2022)	**Expert AGI** not yet achieved
Level 4: Virtuoso *at least 99th percentile of skilled adults*	**Virtuoso Narrow AI** Deep Blue (Campbell et al., 2002), AlphaGo (Silver et al., 2016; 2017)	**Virtuoso AGI** not yet achieved
Level 5: Superhuman *outperforms 100% of humans*	**Superhuman Narrow AI** AlphaFold (Jumper et al., 2021; Varadi et al., 2021), AlphaZero (Silver et al., 2018), StockFish (Stockfish, 2023)	**Artificial Superintelligence (ASI)** not yet achieved

Thus, even though we may currently only have Level 1 “Emerging” AI and AGI, this level of AI is more than sufficient to prompt deep reflection on the purpose and place these AI tools have in the T&L space. In the same way that the user can instrumentally or intrinsically value specific capabilities such as instrument playing or foreign language learning, AI tools are poised to disturb, decouple, and reveal which human capacities are ultimately instrumental or intrinsically valued by teachers and students alike. As we consider how AI can be applied to enhance higher education, we will need to simultaneously contend with how AI transforms what it means to teach and to learn in the 21st century.

To this end, in this section, we present some data and case studies from qualitative interviews of both teachers and students to understand better how AI is being used at HKU and other universities in and out of the classroom. We found that the integration of GenAI into higher education presents a dual-edged reality. A prominent theme across both perspectives was the potential for enhanced efficiency and productivity, with AI tools facilitating tasks ranging from content generation and summarisation for instructors to streamlining research and drafting for students. AI is perceived to be a valuable support for learning and comprehension, offering students personalised explanations and clarifying complex concepts while aiding instructors in structuring materials more effectively.

However, significant concerns arose regarding the accuracy and reliability of AI outputs, particularly in technical disciplines, necessitating critical evaluation and fact-checking. Both students and teachers also express apprehension about the risks of over-reliance, which could potentially impede the development of essential critical thinking, problem-solving, and foundational skills. While AI can foster engagement and motivation through interactive learning experiences, there is a parallel concern that the ease of access to AI-generated solutions might diminish deeper engagement with course content and independent thought, with unintended and unforeseen consequences.

Ultimately, the integration of AI in higher education requires ongoing consideration of its benefits alongside its limitations and ethical implications to ensure it serves as a tool to augment, rather than supplant, effective pedagogical practices and genuine student learning.

2. Common themes from both teachers and students

2.1. Verifying AI-generated output

By far, the most common concern from both educators and students was the recognition that AI tools, as they currently are in 2025, are still prone to inaccuracies, hallucinations, and can provide unreliable information some percentage of the time. This can be especially salient in complex or specialised domains like mathematics, law, and physics. The inherent unreliability necessitates critical evaluation and verification of AI-generated content.

> *"For the last period, my last teaching experience was in last year and back then the AI hasn't been very good and I was using the GPT 4.0 and it was not, usually it gives, it gives me answer, fluent answers, but sometimes it hallucinates and the results, sometimes the logics are not, just not right, so I need to modify myself." (Teacher 1289)*

A common strategy for verifying AI output was to cross-reference the information provided by AI with established, reliable external sources. These include textbooks, lecture notes, academic journals, reputable websites, Google Scholar, and general web searches, suggesting that AI may not necessarily replace knowledge generation in areas where such knowledge cannot be immediately or cheaply verified (e.g., coding snippets).

> *"So I will cross check with textbooks and also journal articles and existing materials from our faculty to make sure the content suggested or generated is really accurate." (Teacher 1741)*

> *"Sometimes I verify what AI has generated against the course materials, such as the readings. And since for my coursework... requires a higher level of higher detail and also legal analysis. Most of the time I will still go back to Google Scholars or HKUL to double check on the information..." (Student 599)*

In classroom contexts, teachers and students alike still often rely on their own existing knowledge, subject matter expertise, critical thinking skills, and intuition to assess the quality and accuracy of AI-generated content, especially for more recent developments in their field of expertise. They evaluate whether the output aligns with their understanding, course requirements, or professional standards. However, there is also recognition that these flaws in AI output are rapidly being reduced as AI improves over time.

> *"I evaluate AI-generated materials informally. I rely on student feedback and discussions... My assessment is mostly based on informal feedback from students and teachers." (Teacher 1563)*

> *"I evaluate AI-generated output based on my foundational knowledge. If I don't have a strong understanding of the topic, it's hard for me to judge the quality...*

I have noticed that AI does well with older concepts but struggles with contemporary ones" (Student 851)

3. Teaching and teacher perspectives

3.1. How teachers use GenAI

Based on teacher interviews, the most prevalent application revolves around generating course content, materials and structuring. Educators leverage GenAI as an efficient tool to create initial drafts of various teaching materials. This includes formulating course outlines, writing scripts (e.g., for videos), developing assignment questions and associated rubrics, crafting scenarios for case studies, and generating relevant text or images for presentations like PowerPoints. Teachers find it helpful in establishing frameworks and structures for their teaching content, helping them organise materials in a way that is easier for students to understand. For instance, AI can help structure a presentation or suggest key points to cover. While AI-generated content often requires modification to fit specific course needs, it significantly boosts efficiency in course design.

The second major use case identified is course content modification and refinement. Teachers employ GenAI not just to create new content, but also to enhance existing materials. This involves tasks such as improving the language and tone of communications (e.g., making feedback friendlier or emails more polished, correcting grammar, paraphrasing text, summarising complex information, and augmenting presentations with better organisation or visuals. AI can help refine PowerPoints by suggesting missed topics or areas for consolidation.

Beyond content creation and refinement, teachers also utilise GenAI to generate ideas and brainstorm. It serves as a source of inspiration for developing interactive classroom activities, formulating engaging discussion questions, designing icebreakers, or exploring different perspectives on a subject. AI can suggest novel examples or approaches that educators might not have considered themselves. Furthermore, GenAI is employed for analysis and summarisation, helping teachers quickly grasp key points from readings or webpages, or analyse concepts to better prepare for teaching. Finally, some educators have used AI to assist with generating feedback and assessment, such as drafting preliminary feedback based on rubrics or utilising AI-enhanced tools like Gradescope to streamline the marking process, particularly for tasks like identifying common answers or grading multiple-choice questions.

While less common, some unique applications emerged. A few educators are involved in Developing Custom AI Tools, such as chatbots tailored to specific courses (e.g., “Navi” or a tool for social work skills) designed to guide students or provide course-specific information. Notably, some projects even Involve Students in AI Tool Development, allowing them to contribute to the knowledge base or annotation, thereby learning about both AI and the subject matter. One teacher mentioned using AI for personal curriculum analysis and evaluating university courses relative to career paths.

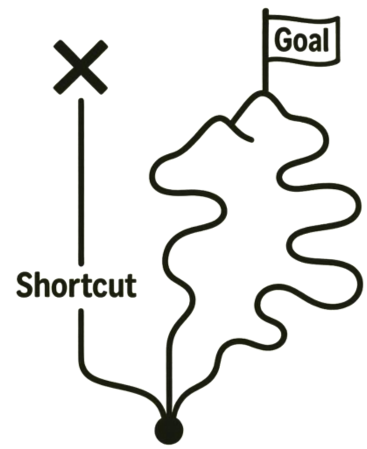
Table 3. Teacher AI Use Cases

Common Use Cases	Description & Examples
Content/Material Generation & Structuring	Drafting course outlines, scripts, assignment questions, presentation frameworks, rubrics, case study scenarios, generate images/text for slides.
Content Modification & Refinement	Improve language/tone, polish grammar, paraphrase, summarise, augment existing presentations, and refine slides.
Idea Generation & Brainstorming	Develop ideas for course content, interactive activities, discussion questions, icebreakers, explore different perspectives.
Analysis & Summarization	Summarise readings, extract key points from web pages, analyse concepts for teaching preparation.
Generating Feedback & Assessment	Draft preliminary feedback using AI-assisted tools (e.g., Gradescope) for marking assistance.
Less Common Use Cases	**Description & Examples**
Developing Custom AI Tools	Create bespoke chatbots or AI systems for specific pedagogical needs (e.g., course guidance, skill practice).
Involving Students in AI Development	Engage students in the creation/training process of custom AI tools.
Personal Curriculum Analysis	Evaluate course relevance for career paths.

3.2. Teacher Perceptions on Student Learning & Engagement

Teachers currently hold varied views on the effectiveness of GenAI in enhancing student learning outcomes and engagement, often relying on informal observations rather than formal metrics. One prominent theme emerging from their responses is the potential for increasing motivation, student engagement, and course accessibility. Several educators noted that students seem more motivated and find learning more interesting when AI tools are involved. The interactive nature of AI, its ability to patiently answer numerous questions (even “strange” ones), and its capacity to provide detailed, line-by-line explanations for complex concepts can make learning feel more achievable and user-friendly. Features like simulations or virtual labs offered by some AI platforms can further stimulate engagement. This perceived boost in confidence and interest is seen as a positive influence.

However, this optimism is tempered by significant concerns about reduced critical thinking and skill development. One major worry among educators is that excessive reliance on GenAI could impede students’ ability to develop fundamental academic skills. They fear that students might use AI as a shortcut, bypassing the necessary cognitive effort required for critical thinking, effective writing, deep reading comprehension, problem-solving, and even memorisation. There is a concern that using these tools too early, before foundational skills are established, could be detrimental. Some teachers speculate that for students primarily focused on obtaining a degree rather than deep learning, AI might enable them to “stop thinking”. Seeing this major concern, we thus offer some suggestions in the section “Suggested Pedagogical Approaches” on how to incorporate AI into the classroom that would help enable and enhance student critical thinking skills.



Relatedly, teachers distinguish between efficiency and surface-level improvements versus deeper understanding of the course material. While acknowledging that AI can efficiently enhance the surface quality of student work, such as improving language fluency or summarisation speed, educators question whether this translates to genuine comprehension or mastery of the subject matter. Polished language generated by AI doesn’t necessarily equate to better content quality, critical analysis, or proper academic practices like referencing. There is a concern that students might focus on merely completing assignments quickly using AI, potentially becoming less creative or innovative and lacking the motivation to learn beyond the task requirements. However, some see potential if AI provides a framework, encouraging students to engage in deeper discussions.

Finally, teachers emphasise that the impact of GenAI in the classroom can be varied. Its effect on learning and engagement is not uniform; it depends heavily on the individual student’s motivation, how they choose to use the tool, and the specific task. Some students might be motivated to excel further using AI, while others might become complacent or “lazy”. This could potentially create a divide, with some students using AI to deepen their learning and others using it to minimise effort. Evaluating these varied impacts often relies on informal methods, such as observing classroom engagement levels or gathering anecdotal feedback through discussions.

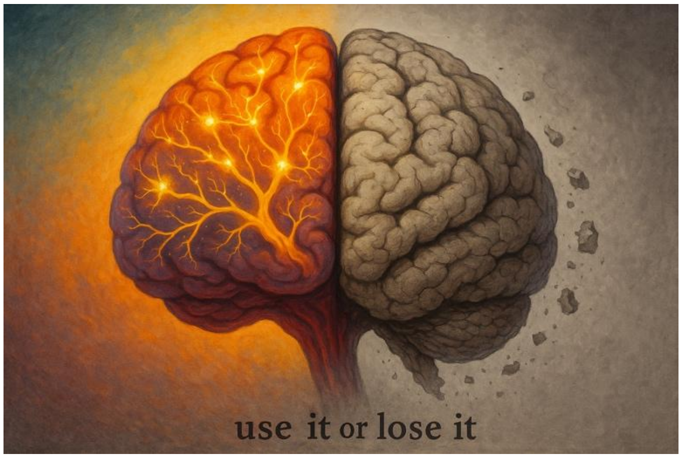
Table 4. Teacher Perceptions on Student Learning & Engagement

Positive Impacts / Observations	Concerns / Negative Impacts
Increased Motivation & Engagement: Students find AI interesting, engaging, fun.	Reduced Critical Thinking: Over-reliance may hinder critical analysis skills.
Enhanced Accessibility: AI patiently answers questions, explains complex ideas simply.	Hindered Skill Development: May impede writing, reading comprehension, problem-solving skills.
Boosted Confidence: Makes tasks feel more achievable.	Superficial Learning: Polished output doesn't guarantee deep understanding or content quality.
Interactive Learning: Tools like simulations can make learning enjoyable.	Potential for Reduced Effort/Creativity: Students might use AI as a shortcut, becoming less innovative.
Efficiency: Can help students produce work faster (Implied, related to teacher efficiency).	Over-reliance: Students might become too dependent, lacking skills needed post-graduation.
Varied Impact: Can motivate some students to excel further.	Widening Gap: Potential divide between students using AI for deep learning vs. minimal effort.
Informal Evaluation: Engagement in class used as a gauge.	Informal Evaluation: Lack of formal metrics to assess true learning impact.

3.3. Teacher Perspectives on Limitations and Ethics

Alongside the potential benefits, teachers identify significant limitations and ethical challenges associated with using GenAI technology in higher education. A primary concern centers on the accuracy, reliability and quality of AI output. Educators frequently encounter issues with AI-generated content, including "hallucinations" – the creation of false information or non-existent references. Factual inaccuracies, superficial treatment of complex topics, and inconsistent outputs are common problems. A particular challenge is that AI models often present information with a high degree of confidence, even when the information is incorrect, stemming from their probabilistic nature. This necessitates significant human oversight, critical evaluation, and fact-checking of AI outputs.

The impact on student learning and skill development discussed previously as a concern regarding effectiveness, is also framed as a fundamental limitation. Teachers worry that the availability and ease of use of GenAI tools actively hinder the development of essential academic



competencies. Skills like critical thinking, analytical writing, thorough reading, independent problem-solving, and even memorisation may atrophy if students consistently rely on AI instead of engaging in the cognitive processes required to build these skills. There's a specific concern about the long-term consequences, questioning whether students will possess the necessary skills for the workplace where AI support might not be available or appropriate.

GenAI poses substantial academic integrity and assessment challenges. The ease with which students can generate text or other content using AI makes it extremely difficult for educators to detect plagiarism or unauthorized use. This complicates the fair assessment of genuine student effort, understanding, and capabilities. Teachers struggle to determine how much of a submitted work is the student's own versus AI-generated, raising questions about how to design assessments that are both meaningful and resistant to AI misuse. The potential for students to become over-reliant on these tools, especially under pressure or uncertainty, further exacerbates these assessment concerns. Some teachers view AI primarily as a threat to students investing the necessary time and effort in their learning.

Finally, there are also significant concerns over the equity, and fairness in the availability of AI. Disparities in access to GenAI tools – such as free versus paid versions, or geographical restrictions – can create an uneven playing field. Students with better access or stronger technical skills ("prompt engineering") may gain an unfair advantage over peers who lack such resources or knowledge. This presents an equity dilemma: mandating AI use could disadvantage those with limited access, while forbidding it might disadvantage students who could benefit ethically. Furthermore, ethical issues include potential privacy and confidentiality breaches if sensitive student or course information is uploaded to external AI platforms. The inherent biases potentially embedded within AI models also raise ethical flags, although not explicitly detailed in these excerpts. The comparison is made that while paying a human editor might be considered ethical, the equity implications of AI access present a new challenge.

Table 5. Teacher Concerns: Limitations and Ethical Issues of GenAI in Higher Education

Category	Specific Concerns & Limitations
Accuracy, Reliability, & Quality	Hallucinations (fake information or references), factual inaccuracies, superficiality, lack of depth, inconsistency, confident presentation of incorrect information.
Impact on Student Learning & Skills	Hinders development of critical thinking, writing, reading comprehension, problem-solving, memorisation; risk of skill atrophy; long-term workplace readiness.
Academic Integrity & Assessment	Difficulty detecting AI use/plagiarism, challenging to evaluate genuine student effort/understanding, need for AI-resistant assessment design, over-reliance.
Equity, Fairness, & Access	Disparities due to paid tools vs. free, geographical restrictions, varying technical skills (digital divide), potential for unfair advantages.
Privacy & Confidentiality	Risk of sensitive data misuse if uploaded to AI platforms (e.g., placement details).
Ethical Considerations	Potential for inherent bias in AI models (Implied), ethical dilemmas regarding equitable use vs. non-use policies.

4. Learning and student perspectives

4.1. Student Use of GenAI

Student use of GenAI tools encompassing text and image synthesis capabilities (though student feedback primarily focused on text/code generation) is increasingly integrated into the academic lives of higher education students. Based on student self-reporting, these technologies are employed across a diverse spectrum of learning activities, fundamentally altering traditional study workflows. One of the most prominent uses is enhancing comprehension and clarifying complex subject matter. Students frequently turn to AI assistants like ChatGPT, Poe, DeepSeek, Claude, and Perplexity to break down complex concepts encountered in lectures or readings, simplify dense academic language found in textbooks or legal judgments, summarise lengthy materials, and receive alternative explanations when lecture content is unclear or missed. This function allows students to personalise their learning pace and address knowledge gaps quickly, acting as an on-demand resource for clarification.

Beyond conceptual understanding, GenAI serves practical roles in assignment completion and skill development. Students use these tools to generate initial ideas and create structural outlines for essays and reports, helping them overcome writer's block and providing a starting point for written tasks. In technical fields like computer science, engineering, mathematics, and physics, AI is used for coding assistance, debugging code, and attempting to solve homework problems, although often with noted limitations in accuracy. Furthermore, students leverage AI to refine their written work through proofreading for grammatical and syntax errors and improving sentence structure and effectiveness.

Other reported applications include creating personalised study plans, generating practice exercises for language learning (e.g., French conjugations), translating content, conducting preliminary research or finding relevant sources, and even learning about external fields like finance and the stock market. The breadth of these applications underscores GenAI's role as a versatile assistant in various facets of student academic life.

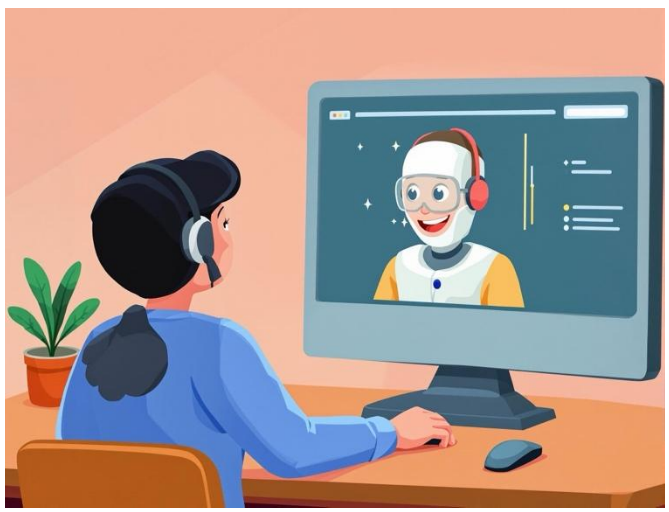
Table 6. Key Ways Students Use GenAI in Higher Education

Category of Use	Description
Concept Clarification & Learning	Simplifying complex topics, explaining jargon, summarising readings/lectures, answering specific questions.
Assignment Assistance	Generating ideas, outlining essays/reports, drafting sections (e.g., introductions), coding help.
Problem Solving	Attempting homework problems, particularly in quantitative fields like math and physics.

Writing & Language Support	Proofreading, grammar/syntax checks, rephrasing sentences, improving language skills, translation.
Study & Research Aid	Creating study plans, generating practice exercises, preliminary research, finding sources.

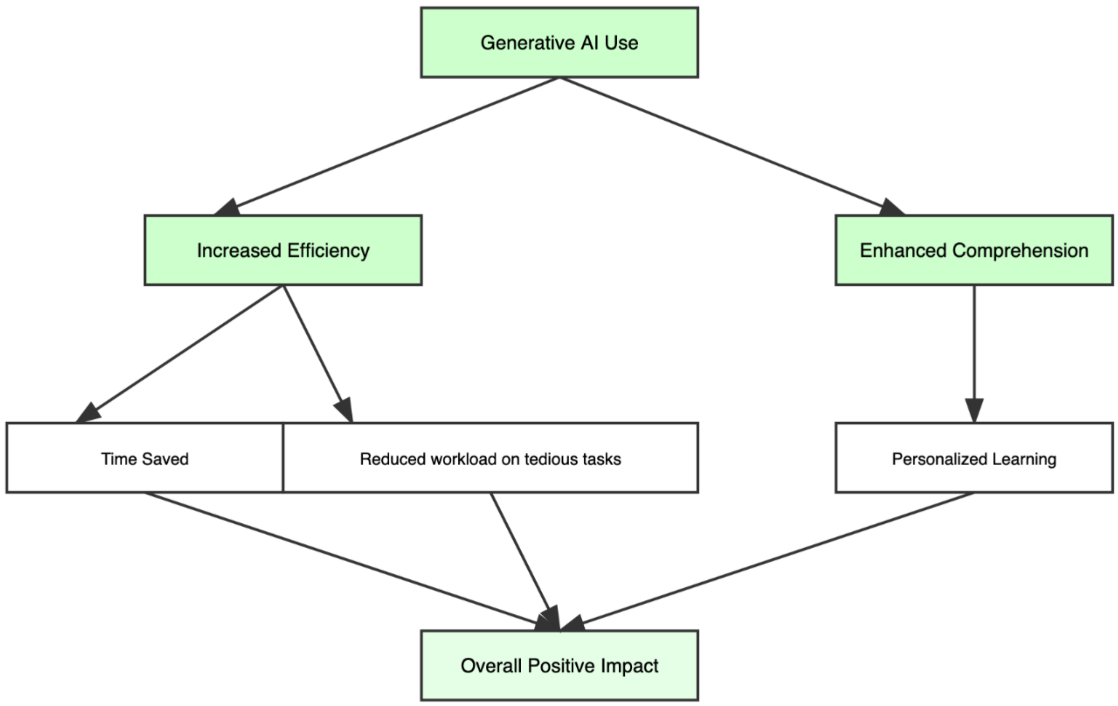
4.2. Perceived Effectiveness and Benefits for Students

Students largely perceive GenAI tools as effective instruments for enhancing their academic productivity and supporting their learning processes, although this perception is often tempered by concerns discussed later. The most frequently cited benefit is a significant improvement in efficiency and time management. Students report that AI drastically reduces the time spent on laborious tasks such as summarising extensive readings or lecture notes, conducting initial literature searches, drafting emails, outlining papers, generating code snippets, and translating documents. This time-saving aspect allows students to allocate more effort towards understanding core concepts, critical analysis, or managing other academic and personal commitments. The ability to quickly brainstorm ideas or structure assignments also helps overcome procrastination and writer's block, making the process of starting tasks less daunting.



Beyond workflow optimisation, GenAI is highly valued for its role in enhancing comprehension and providing personalised learning support. Beyond the aforementioned use in helping students break down complex theories or processes into simpler, more digestible terms, students often request explanations tailored to their current level of understanding, which may range from basic to advanced. This function not only serves as a readily accessible, supplementary learning resource, akin to an "instant tutor", capable of clarifying doubts immediately without the potential delay or formality of approaching instructors or teaching assistants, but also provides a reported key psychological benefit of avoiding human judgement and potential shame for asking questions due to the non-judgmental nature of AI. Students feel more comfortable asking rudimentary or "dumb" questions that they might hesitate to pose to peers or faculty, thereby facilitating a more open and less inhibited learning process. Learning becomes more self-directed and students can explore topics at their own pace and delve deeper into areas missed or poorly understood during formal instruction. Teachers may consider how GenAI may change the relationship and dynamic between student and teacher in the traditional setting.

Figure 1. Student-Reported Benefits of Using GenAI



4.3. Student Perspective on Limitations, Risks, and Ethics

Despite their recognised benefits, students express significant concerns and identify several limitations regarding the use of GenAI in higher education. A primary and recurring issue is the accuracy and reliability of AI-generated content. Students frequently report encountering factual errors, incorrect calculations, flawed logic, and fabricated information, particularly in technical subjects like mathematics, physics, statistics, and coding. The generation of non-existent sources or “hallucinated” references is a common frustration, necessitating time-consuming manual verification against reliable academic databases, textbooks, or lecture notes. This unreliability means that at least some students recognise that they cannot implicitly trust AI outputs and must engage in critical evaluation and fact-checking, sometimes negating the time saved. AI outputs can also be superficial, lack necessary context, or be outdated, limiting their utility for nuanced or specialised tasks.

Another major area of concern revolves around the potential for over-reliance on GenAI and its detrimental impact on learning and skill development. Students worry about becoming “lazy”, losing motivation to think independently, and experiencing a decline in critical thinking, problem-solving, and analytical abilities. There’s apprehension that excessive use might hinder the development of fundamental skills, such as writing proficiency, coding from scratch, or deep research, as students might passively follow AI suggestions without achieving genuine understanding. Some students noted that reduced reliance could paradoxically lead to better academic outcomes, suggesting that AI assistance doesn't always equate to deeper learning.

Ethical considerations are also prominent in student perspectives. Concerns about equity and fairness arise from the potential disparity between students who can afford premium AI models versus those using free versions, and varying levels of AI literacy create disadvantages. Accessibility issues, particularly in regions like the Global South, further exacerbate these equity concerns. Academic integrity is another critical point, with students worried about the misuse of AI for plagiarism, the potential for fellow students to submit unearned work, and the fairness and accuracy of AI detection tools potentially penalising original work. The lack of clear, consistent university policies on acceptable AI use creates confusion and anxiety. Finally, privacy and data confidentiality emerge as concerns, with students questioning how their input data, potentially including sensitive research information, is used by AI platforms.

Table 7. Summary of Student Concerns Regarding GenAI in Higher Education

Concern Category	Description
Accuracy & Reliability	AI generates incorrect information, fake citations, flawed logic, especially in technical fields; requires extensive fact-checking.
Over-reliance & Skill Erosion	Fear of reduced critical thinking, laziness, hindering development of writing, coding, problem-solving skills; superficial understanding.
Equity & Accessibility	Disparities due to paid vs. free tools, varying AI literacy, unequal access across regions/resource levels.
Academic Integrity & Policy	Misuse for plagiarism, unfair advantage, unreliable AI detection, lack of clear institutional guidelines, potential devaluation of degrees.
Privacy & Confidentiality	Concerns about how user prompts and potentially sensitive data are stored andutilised by AI platforms.

In conclusion, student perspectives reveal that while GenAI offers powerful tools for enhancing efficiency and supporting learning in higher education, its adoption is accompanied by significant challenges related to reliability, the potential for deskilling, and complex ethical considerations including fairness, academic integrity, and privacy.

Section III:
Application and Case Studies

1. Overview

This section explores some creative and experimental examples of GenAI integration in teaching and learning. Through practical applications and case studies, we illustrate how GenAI can foster critical thinking, deepen student engagement, and provide immersive and interactive learning experiences. The examples focus on addressing challenges in humanities education, critical thinking development, and speculative futures, while offering actionable strategies for educators.

By examining applications such as AI-enhanced role-play, structured prompt engineering, and interactive debates, as well as case studies like Picturing China – AI vs. the World and Future of Made-in-Hong Kong, we highlight diverse ways GenAI can be used to enrich pedagogy. They not only support creative and analytical processes but also promote collaboration, critical reflection, and the development of AI literacy, equipping students with skills essential for navigating an increasingly AI-driven world.

2. AI- Embodiment and Role-Play in Humanities Ed

The potential integration of AI-powered embodiment and role-play presents a transformative opportunity for humanities education in higher learning environments. By connecting these complementary concepts more effectively, we can illustrate how immersive AI experiences can deepen student engagement with humanities subjects through embodied learning.

AI-enhanced embodiment through role-play offers unprecedented opportunities to revolutionise humanities education by allowing students to physically and emotionally experience different perspectives, periods, and cultural contexts. This embodied approach addresses a fundamental challenge in humanities education—making abstract concepts, historical events, and literary works tangible and personally meaningful to students.

When students engage in AI-facilitated role-play, they develop enhanced empathy and perspective-taking abilities by literally “stepping into the shoes” of historical figures, literary characters, or individuals from different cultures. Unlike passive reading, this embodied learning creates a visceral understanding as students physically enact different perspectives and receive dynamic feedback from sophisticated AI systems that can challenge assumptions and present alternative viewpoints. This embodied engagement helps students recognise their own biases and develop a nuanced understanding of complex social and historical issues.

The immersive learning experiences created through AI-enhanced role-play transform traditionally passive humanities study into active, engaging processes. For example, AI systems can enable students to embody characters from literary works, allowing them to experience the narrative’s emotional landscape through a first-person perspective physically. This embodiment creates “living texts” where students don’t just read about experiences but enact them, making learning more memorable and emotionally impactful.

The exploration of counterfactuals becomes particularly powerful through AI-enhanced embodiment, as students can physically experience alternative historical scenarios or different character decisions. By literally embodying e.g., voice-changing AI or character swap AI, these “what if” situations, students develop a deeper understanding of causality and contingency while strengthening critical thinking skills through active participation rather than abstract consideration.

Practical applications of AI-enhanced embodiment span across humanities disciplines: literature students can physically embody Shakespearean characters, history students can experience decision-making as historical figures, philosophy students can engage bodily in philosophical debates through AI-facilitated role-play, and cultural studies students can participate in cross-cultural simulations that create an embodied understanding of different cultural contexts.

2.1. Case study: GOAT - Who is the Greatest Economist of all Time and Why Does it Matter?" By Tyler Cowen

In an intriguing project by economist Tyler Cowen, the author of "GOAT: Who is the Greatest Economist of all Time and Why Does it Matter?", Cowen employed artificial intelligence (AI) as a research, drafting, and editing assistant, an increasingly commonplace practice. Most interestingly, however, Cowen also made an experimental decision to utilise AI as a pedagogical companion for all readers of what he refers to as a "generative book". EconGOAT GPT-4 (econgoat.ai), an AI chatbot trained explicitly on the book's content, was simultaneously released with the book in the hope of enabling readers to engage with the material dynamically, through role-play. This interactive resource allows readers to pose questions about economists and their theories, seek clarifications, and even engage in debates with Cowen's conclusions. The multifaceted purpose of this integration is to enhance reader engagement, facilitate deeper learning by providing an immediate interactive resource, and explore the potential of AI in augmenting educational experiences. By offering this AI companion, Cowen transforms the traditional reading process into a more interactive dialogue with the subject matter by allowing readers to:

- pose questions to the AI about the economists discussed in the book
- debate Cowen's conclusions or the merits of different economists with the AI
- generate multiple-choice exams on the book's content or even help create an illustrated version of a chapter (Cowen, 2023)

In the teaching and learning context, it would be highly feasible for teachers to implement similar innovative strategies of facilitating deeper engagement with older classic texts, which are likely to be within the training corpus for most GenAI systems, or to create and train their own for use in the classroom.

3. Critical thinking through AI prompt engineering

The integration of Generative Artificial Intelligence (GenAI) in university settings has sparked concerns among educators regarding its potential impact on students' ability to develop critical thinking skills. While GenAI tools have been shown to enhance cognitive skills, including critical thinking, through personalised learning and feedback (Daniel et al., 2025), there is apprehension that reliance on these tools might diminish students' engagement with the source material and their ability to evaluate information critically.

Teachers worry that the ease of generating content with GenAI could lead to a superficial understanding of topics, as students might bypass the rigorous process of scrutinising and synthesising information from diverse sources (Duah & McGivern, 2024; Salter & Bonfield, 2024). Moreover, the lack of clear institutional policies on the ethical use of GenAI in academic work further complicates its integration, potentially leading to academic misconduct and undermining the development of critical thinking (Duah & McGivern, 2024; Hsiao & Tang, 2024). Therefore, while GenAI offers promising educational benefits, universities must implement strategies that ensure these tools are used to complement, rather than replace, traditional learning processes that foster critical thinking and analytical skills (Toscano et al., 2024; Zhou et al., 2024).

Thus, we now offer a few suggestions on how educators could redesign their curriculum in a way that thoughtfully integrates AI to help enhance rather than hinder students' critical thinking skills and learning experience. To facilitate the adoption of using AI purposefully, we outline three specific types of activities first presented in an in-progress manuscript (Tsao, Wong, Mo, 2025 Forthcoming), which educators can adapt for their courses:

1) Critiquing AI responses activity
2) Structured prompt engineering exercises
3) Interactive debating with LLMs

We hope that these activity outlines serve as a springboard of ideas for educators to create and tailor their own methods of using GenAI in classrooms with purpose, amidst a transitional period of institutional pushes for adopting AI without clear accompanying guidelines for how and why.

3.1. Critiquing AI responses

In the first activity on critiquing AI outputs and processes, the exercise encourages students to analyse and critique AI-generated outputs, such as essays, summaries, or creative content. For example, with the release of reasoning models, they could evaluate the relevance, accuracy, and quality of an AI-generated chain of thought in response to a prompt, identifying biases or inaccuracies and proposing improvements. We suggest this activity based on prior interview data showing that the effectiveness of AI output depends in part on the users' ability to ask precise questions and critically interpret AI-generated content (Lawasi et al., 2024).

One interesting exercise teachers could implement in the classroom is an "Expert vs. Novice AI Response Analysis" activity. In this exercise, students compare two AI-generated responses on the same topic—one produced by an AI instructed to simulate expert knowledge and another simulated novice or layperson understanding. The teacher selects a curriculum-relevant topic and generates both responses in advance without labelling which is which. When presenting the activity, the teacher explains that students will analyse both responses to determine which demonstrates expert knowledge and which demonstrates novice understanding based on specific evidence from the texts. For additional challenges, the teacher could even intentionally prompt the AI to simulate and include several common misconceptions or misunderstandings in the novice response, particularly ones that are frequently encountered when studying the chosen topic. The unique advantage of using AI is that teachers can create customised examples with specific flaws to demonstrate to students what these flaws look like in practice, allowing for a targeted learning experience, as theorised in (Stuchlikova & Weis, 2024).

Students in this activity can then work in pairs or small groups to carefully analyse both responses, noting differences in terminology, depth of explanation, quality of examples, acknowledgement of limitations, and overall approach to the topic. For instance, in an environmental sustainability class, students might analyse two responses about climate change solutions, identifying one response's more precise use of terminology supported by specific data points, acknowledging implementation challenges, and discussing complex interconnections between proposed solutions. The other response might use more generalised statements in non-technical language and focus on more linear solutions, unverified supporting evidence, and differences in its presentation.

When comparing the responses, students might contrast these differences in content, such as how the expert responded:

1) Cites specific research studies or refers to particular scholarly frameworks, while the novice response relies more on general knowledge or popular media perspectives.
2) Distinguishes between short-term and long-term implications of various solutions, whereas the novice response treats all timeframes similarly.
3) Addresses regional variations in the applicability of solutions (such as explaining why certain renewable technologies are more effective in specific geographic contexts), while the novice response offers one-size-fits-all recommendations

4) Acknowledge uncertainties and areas of ongoing research, presenting multiple viewpoints within the scientific community, while the novice response presents information with unwarranted certainty.
5) Uses technical vocabulary appropriately (such as distinguishing between “energy density” and “power density” when discussing battery technologies), while the novice might use technical terms incorrectly or interchangeably.
6) Provides quantitative comparisons (like specific percentage reductions in emissions achievable through different approaches), whereas the novice relies on qualitative descriptors like “a lot” or “very effective.”

Students may also observe structural differences—the expert response likely has a more logical organisation that builds on fundamental concepts before addressing complex interactions. In contrast, the novice response might jump between ideas without clear connections. The expert might address potential objections or limitations to their own arguments, demonstrating awareness of counterpoints, while the novice presents information without acknowledging alternative perspectives. Finally, students might note that the expert situates solutions within broader systems (considering economic, political, and social factors that impact implementation). At the same time, the novice treats solutions in isolation without considering real-world contexts.

After their analysis, students can participate in a class discussion where they share their determinations and, crucially, the specific pieces of evidence and reasoning that led them to identify which response demonstrates expertise. The teacher guides this discussion by asking students to articulate what specific language, structure, or content elements revealed greater subject matter knowledge. Following the reveal of the AI-simulated and expert-generated response, students reflect on the markers of expertise they identified and discuss how these evaluation skills could be applied to other information sources they encounter. Finally, the students can even be given the (optional) opportunity to implement this activity themselves through a student-led, active learning and exploration activity, implementing the student-driven innovation theorised in (Dai et al., 2023). As the teacher, students will now have to consider precisely how to prompt the AI in a way that would lead to the AI generating novice vs. expert responses.

This activity can help students develop critical thinking skills by requiring them to move beyond surface-level assessment of information quality to identify specific indicators of expertise and lack of expertise. Rather than simply consuming AI-generated content uncritically, students build discernment skills that allow them to evaluate information quality across various contexts. The initial simulation step is based on the prompting technique known as “Role prompting” (Kuka, 2024) or “Role-play prompting”, which suggests that having AIs adopt specific personas can enhance (Kong et al., 2023) or degrade (Kim et al., 2024) the AI’s reasoning. This tendency can be put to good use in the classroom to elicit instances of better or worse critical thinking and having students compare what makes one output better or worse than the other. Then, by flipping the student-teacher roles, the exercise allows students to see for themselves that AI as a tool can be prompted to generate weaker or stronger arguments for various positions, helping students understand that not all AI-generated content is equally trustworthy or reliable if the full prompt is not shown to them. This approach aligns with media literacy education principles – addressing the unique challenges presented by AI-generated content while simultaneously developing critical thinking rather than replacing it.

3.2. Structured Prompt Engineering

Second, structured prompt engineering activities can be used to develop students' ability to craft precise and effective prompts to guide AI-generated outputs. Prompt engineering is the practice of crafting effective inputs or "prompts" to elicit the best possible outputs from AI models. Prompt engineering exercises can be used to scaffold and incentivise the development of critical thinking competencies through structured engagement with AI systems. This pedagogical approach requires learners to precisely break down and articulate what they are looking for, as well as provide relevant contextual parameters – precisely the skills that underpin critical thinking in alignment with established metacognitive frameworks in the scholarship of teaching and learning. For example, learning to formulate precise questions and provide relevant context develops the following essential skills:

1. Problem decomposition - Breaking down complex queries into clearer, more specific parts
2. Context awareness - Understanding what background information is necessary for meaningful answers
3. Information prioritisation - Deciding which details are essential vs. peripheral
4. Iterative refinement - Learning from vague or incomplete responses to improve future prompts

However, this benefit depends heavily on how students engage with AI. Simply rephrasing questions until getting a desired answer without understanding why specific prompts work better may not develop these skills. Critical thinking comes from questioning the initial output generated by AI models and iteratively testing and analysing why some prompts are more effective than others in order to understand the model's limitations and requirements.

Prompt engineering exercises are especially fruitful as they hit several course outcomes within a single activity. These exercises can be used to simultaneously:

a. deliver useful course content
b. create experiential learning opportunities on critical thinking
c. practice valuable future workplace skills where iterative AI use will be part of their workflow
d. create assessments that are AI-aware or "AI-proof" and scale with improvements in models

Thus, conducting prompt engineering exercises together with learners can be an especially valuable teaching tool for courses whose learning outcomes involve critical thinking. Currently, models respond with higher quality and specificity when given detailed, precise prompts rather than vague or general ones. This pattern is likely to persist even as AI capabilities advance. Teachers can use this property of AI systems to create *experiential* learning opportunities with rapid feedback on how to ask more precise questions, how to provide essential context and how to elicit further clarification from the model when needed. Learners and teachers alike can compare prompts and share techniques to elicit the best responses from AI systems. This method is especially resonant with the student-as-partner model, as learners collaboratively explore AI capabilities and limitations, fostering agency and authentic inquiry. Through iterative prompt refinement, learners develop metacognitive awareness while practising real-world applications of critical analysis that will be valuable in the future workplace.

For instance, one example of a structured prompt engineering exercise could be to have students work in pairs to explore how different approaches to prompt writing affect AI responses. The activity could begin with an introduction to prompt engineering fundamentals, where the teacher demonstrates how slight variations in wording can produce significantly different results from AI systems. This could look like having students start with a basic prompt like *“Tell me about how AI affects jobs and the economy”,* then encouraging students to think of ways to ask for more specific versions such as requesting *“Explain automation displacement, productivity enhancement, and new job creation as three key economic impacts of AI technologies, including specific industry examples and statistical evidence for each.”* Finally, for the third prompt, teachers could guide students into creating a contextually rich request that establishes a clear purpose, audience, and framework for the AI response. The third prompt should be the most detailed, with specific considerations relevant to the field or topic in question, such as:

“I’m looking for a presentation to local business leaders and workforce development officials about preparing our community for AI-driven economic changes. Please analyse how AI is likely to affect employment in manufacturing, healthcare, and service industries over the next decade. For each sector, identify which specific job categories face displacement risks, which might be augmented rather than replaced, and what new roles might emerge. Include perspectives from economists with optimistic views on AI job creation, labour experts concerned about structural unemployment, and industry analysts focused on workforce transition challenges. Also, it will address how different demographic groups (varying by education level, age, and geographic location) might experience these changes differently and what policy approaches could address potential inequities. I need this analysis to help stakeholders develop balanced, forward-looking workforce strategies.”

In this economics/business example, the third prompt not only specifies industries to analyse but also establishes multiple analytical frameworks, stakeholder perspectives, and a clear applied purpose that requires nuanced treatment of a complex socioeconomic issue. Most importantly, in order to create the third prompt, students need to have developed critical thinking skills to identify what sorts of issues, perspectives, and different types of unemployment could exist, what relevant demographic variables should be considered etc., demonstrating the need for domain-specific background knowledge in their asking of the appropriate questions. Teachers can change the number of prompting edit cycles as needed while students submit each prompt to the same AI systems, recording the responses verbatim. This allows them to observe differences between their varying prompts within the same AI. Teachers can also choose to have students submit the same prompts to different AI systems, allowing students to see for themselves how identical prompts might differ across different AI models, which could prompt discussion on how to reconcile any differences that occur between models.

During the analysis phase in between each prompt edit, students can also examine which prompts produced the most useful or accurate information, identify specific language elements that influenced the responses, note any biases or limitations in the AI outputs, and consider how they might improve their prompt strategies. Based on these insights, they then craft and test improved prompts with both AI systems. The exercise can then conclude with small group discussions where students share key discoveries and reflect on effective prompt strategies. For assessment, students could submit their original and refined prompts, their comparative analysis of the AI responses, and a reflection connecting prompt engineering to broader critical thinking skills. Throughout this process, students engage in defining goals with precision, applying background knowledge, providing context for abstract concepts, breaking down problems systematically, evaluating outputs critically, and making adjustments based on the output results—all core components of critical thinking that receive immediate feedback within the AI interaction environment. The benefit of such an assessment model also means that the assessment is automatically AI-aware, as the assessment is not only on assessing the quality of one AI-generated output but rather emphasises assessing the student's ability to process and refine input prompt in addition to the AI-generated outputs.

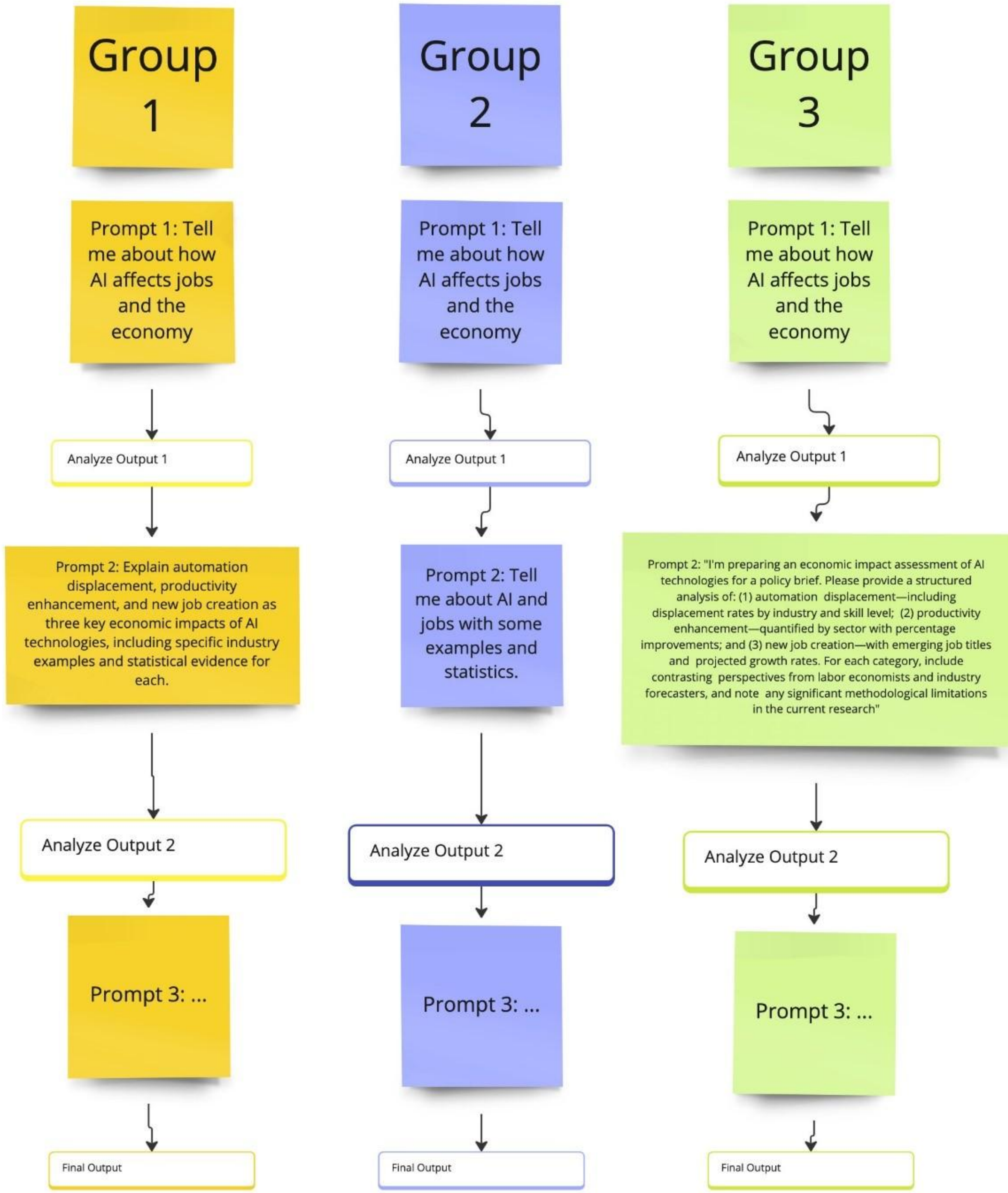


Figure 2. Sample prompt engineering structured exercise flow with 3 groups showing different levels of prompting.

Additionally, this prompt engineering approach naturally establishes AI-aware assessment benchmarks, as AI-generated output from generic or baseline prompts can serve as exemplars against which to evaluate higher-order thinking and more sophisticated work products. For example, teachers and students can generate an initial “baseline” essay using a minimal prompt from multiple popular AI models together, analyse the areas where the AI-generated essays do or do not reach certain assessment criteria, and then use the baseline essay as an exemplar of a minimally acceptable standard essay. Students can then be assessed on how much they can improve upon the baseline AI-generated essay using various prompting techniques or by writing one themselves. This method has the advantage of inherently accommodating and growing with the quality of AI systems, as assessment is based on improvement from the baseline AI-generated essay of a given model quality.

However, in order to use this approach, teachers must ensure that all students are able to access the same AI model(s). At a minimum, the AI system used to generate the baseline essay should be accessible to all students in the course. Teachers can also specify in the assessment rubric that the quality of the prompt and critical thinking skills demonstrated in prompt engineering will be part of the assessment rather than focusing solely on the final “output”. In this way, engineering-based prompt evaluations may actually provide a novel chance for teachers to incorporate direct assessment of the quality of students’ critical thinking chains of thought instead of solely using the proxy of essay output quality.

3.3. Interactive debating with LLMs

Lastly, interactive comparisons and debates with AIs have students analyse the AI's ability to make persuasive arguments on a topic from two opposing perspectives and then critique the quality of reasoning and evidence presented. Alternatively, students could engage in mock debates with AIs, asking targeted questions to challenge the AI's arguments and identify gaps in its logic. These exercises encourage students to think critically about argumentation, reasoning, and the limitations of AI-generated content. We suggest this activity based on prior research showing that even non-experts can assess the correctness of AI models when the assessment is conducted via a debate format (Khan et al., 2024).

One example of an interactive debate activity could be having the students work in small groups to examine how artificial intelligence constructs arguments and the limitations of AI reasoning. Begin by selecting a debatable topic relevant to the curriculum (for example, “Should social media companies be responsible for regulating misinformation?” or “Is space exploration a worthwhile investment of public resources?”).

First, students (in groups or by themselves) can prompt an AI to generate a persuasive argument supporting one position on the topic, then request the same AI to argue the opposing viewpoint with equal conviction. Students should save both responses for analysis. They will then evaluate the arguments using a critical analysis framework, examining the quality of reasoning, types of evidence cited, logical fallacies present, and persuasive techniques employed. Students should specifically identify where the AI makes unsupported claims, overlooks important counterarguments, or demonstrates inconsistencies in reasoning. In the second phase, groups will engage in a live “interrogation” of the AI, asking follow-up questions designed to test the depth of the AI's understanding and expose gaps in its reasoning. For instance, students might ask the AI to provide specific examples or statistics supporting its claims, to respond to counterarguments not addressed in its initial response, or to clarify apparent contradictions. After completing both phases, the class will reconvene to share insights about the strengths and limitations of AI-generated argumentation, discussing what this reveals about critical thinking skills that remain uniquely human.

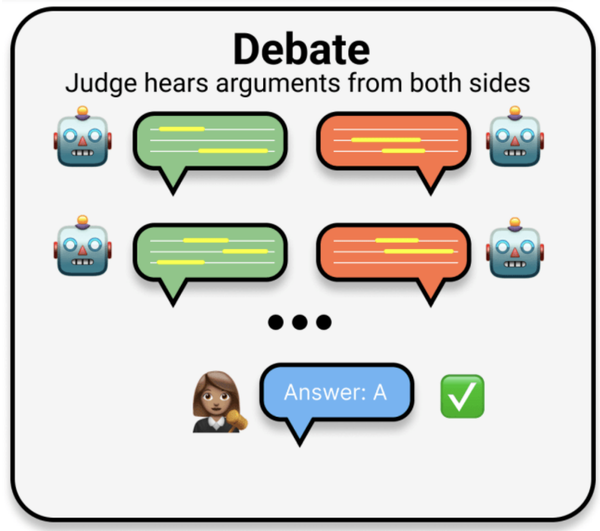


Figure 3. Figure from Khan et al. (2024) illustrating a debate between two LLMs and a human judge.

The activity concludes with students reflecting on how this exercise has enhanced their ability to evaluate arguments they encounter in various media. This debate activity can also be adjusted in multiple ways for students at different stages of academic development. For first-year students or those new to critical analysis, teachers could provide a structured evaluation rubric with specific criteria for assessing arguments (clarity, evidence, logic, etc.) and simplified prompts for the AI interrogation phase. This scaffolding helps students focus their analysis and build confidence in critical evaluation before moving to more open-ended critique. For discipline-specific applications, teachers could frame the debate topics within their subject area—science teachers might explore scientific controversies like gene editing ethics, history teachers could examine historical interpretations of events, and literature teachers might focus on character motivations or thematic interpretations in texts. Teachers can do this by creating customised chatbots for this activity and uploading specific material (e.g., books, articles, stories, etc.) to custom chatbots for debate in order to tailor this activity to the curriculum.

Finally, to integrate multimedia and communication skills, teachers could have students create short presentations or infographics that visually represent the strengths and weaknesses they identified in the AI's arguments, including creating a "reasoning map" that highlights logical connections and gaps. Students could then present these analyses to the class, practising their own persuasive communication skills while deepening their understanding of effective argumentation.

Together, these three activities highlight how GenAI can be used critically in the classroom. The benefits of outlining these activities are to provide concrete examples of how GenAI can be used to give students real-time feedback and personalised output on individual learning paths through active engagement with AI output. By explicitly articulating how GenAI is being used to enhance critical thinking, we hope to provide a starting point for the use of GenAI in promoting the *awareness* of critical thinking itself, which was found to have no significant impact in prior research (Jia & Tu, 2024).

3.4. Case Study: Claude for Education

In the recently released Anthropic Report on Claude for Education (*Anthropic Education Report*, 2025), Anthropic identified four interaction styles most commonly used by students when interacting with Claude.ai. As mentioned in section 3, AI can be used not just as a provider of solutions or answers. The sophisticated AI user (teacher or student) can find any number of ways to incorporate AI collaboratively for iterative improvement.

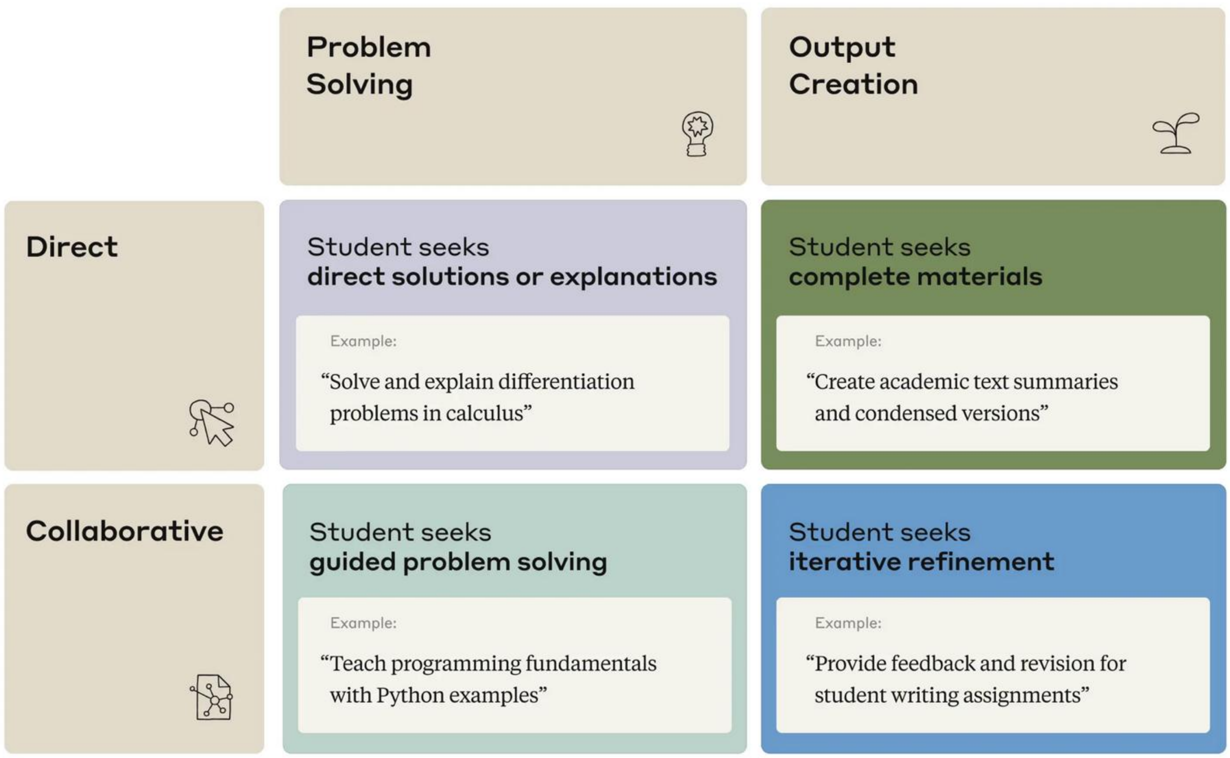


Figure 4. From Anthropic's taxonomy for student-AI conversations, along with sample conversation topics based on those surfaced by Clio (Anthropic Education Report, 2025).

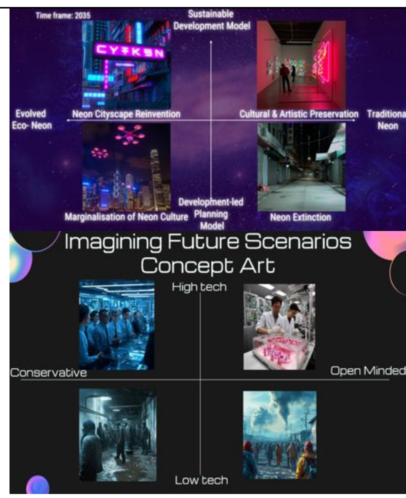
4. Case Study: Picturing China – AI vs. the World

Overview	Picturing China – AI vs. the World is an educational game developed by a PhD candidate in the PPA Department. This short-term classroom activity uses AI generative tools like Claude and Flux AI to foster critical thinking about global perceptions of China by comparing AI-generated and crowd-sourced visuals. It is primarily suited for social sciences, humanities, and media studies.
AI Use	• Utilises LLM Claude to generate descriptive prompts about global perceptions of China. • Employs Flux AI to create images based on AI and crowd-sourced inputs. • Facilitates student discussion to refine crowd-sourced prompts. • Encourages critical analysis of AI-generated versus human-driven outputs. • Develops students' understanding of AI tools as visualisation and critical thinking aids.
Instructions	The activity spans 45–60 minutes and involves four steps: 1. **Introduction** (15 minutes): Educators explain goals, AI tools, and processes 2. **AI Image Generation** (15 minutes): Claude generates a descriptive prompt based on *"Describe a visual scene that represents how the world views China today. Include specific elements, colors, and moods."* and the given prompt is input to Flux AI to create a visual images. 3. **Crowd-Sourced Image Generation** (20 minutes): Students contribute ideas for an alternative prompt, which is refined collectively and used to generate a second image. ▪ Questions: "What elements do you think should be in an image showing how the world views China? Think of symbols, landmarks, stereotypes, or modern features." ▪ Write down on paper some descriptions summarising in a few words how they perceive the picture generated by the AI model. 4. **Discussion and Reflection** (20 minutes): Students compare the two images, analyse biases, and complete reflective tasks to summarise their learning. ▪ *How* does *the AI's view of China reflect its training data? Is it more idealised" or "generic"? What do the crowd's choices say about common global perceptions or biases about China? How might these images differ if created by people in China versus outside it?".*
Impact on Learning	• Enhances critical thinking by comparing AI and human perspectives. • Builds awareness of AI's strengths and limitations in cultural interpretation. • Promotes collaborative learning through crowd-sourced input. • Introduces students to the practical use of AI tools for visual analysis. • Encourages discussions on biases, media influence, and global perceptions.

5. Case Study: Future of Made-in-Hong Kong

By Common Core Office

<table>
<tr><td>Overview</td><td colspan="2">Research project between Arup Foresight’s Team and HKU students exploring the future of industrialisation in Hong Kong by engaging in foresight and futures thinking to contribute to strategic planning and policy thinking.</td></tr>
<tr><td>AI Use</td><td colspan="2">LLMs and image-generation (DALL-E, Midjourney) applications for:

• Future Scenarios: Brainstorm and extrapolate different future scenarios (e.g., high-tech smart manufacturing hubs, sustainability-focused industries, etc.).
• Visualising Speculative Artefacts: Visual conceptual designs of future factories, industrial products, or urban spaces in Hong Kong.
• Storytelling: Support the ideation and writing of compelling future narratives.</td></tr>
<tr><td>Impact on Learning</td><td colspan="2">• AI literacy for supporting creative processes ideation and visual representations (otherwise technically difficult).
• Bring speculative concepts to life, creating vivid, immersive visualisations.
• Assist students in written and visual communication of ideas.
• Students learn to critically assess the quality and relevance of AI outputs.</td></tr>
<tr><td>
</td><td>
</td><td>
</td></tr>
<tr><td>In-person brainstorming and ideation for the reindustrialisation of Hong Kong.</td><td>Extrapolation of future scenarios using GenAI LLM to write stories and image generation.</td><td>Visualising speculative urban spaces, including prototyping of artefacts.</td></tr>
</table>

References

Anthropic Education Report: How University Students Use Claude. (2025, April 8). https://www.anthropic.com/news/anthropic-education-report-how-university-students-use-claude

Cowen, T. (2023, October 23). **GOAT: Who is the Greatest Economist of all Time, and Why Does it Matter?**. Marginal REVOLUTION. https://marginalrevolution.com/marginalrevolution/2023/10/goat-who-is-the-greatest-economist-of-all-time-and-why-does-it-matter.html

Dai, Y., Liu, A., & Lim, C. P. (2023). Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education. *Procedia CIRP*, *119*, 84–90. https://doi.org/10.1016/j.procir.2023.05.002

Daniel, K., Msambwa, M. M., & Wen, Z. (2025). Can Generative AI Revolutionise Academic Skills Development in Higher Education? A Systematic Literature Review. *European Journal of Education*, *60*(1), e70036. https://doi.org/10.1111/ejed.70036

Duah, J. E., & McGivern, P. (2024). How generative artificial intelligence has blurred notions of authorial identity and academic norms in higher education, necessitating clear university usage policies. *The International Journal of Information and Learning Technology*, *41*(2), 180–193. https://doi.org/10.1108/IJILT-11-2023-0213

Hsiao, C.-H., & Tang, K.-Y. (2024). Beyond acceptance: An empirical investigation of technological, ethical, social, and individual determinants of GenAI-supported learning in higher education. *Education and Information Technologies*, 1–26. https://doi.org/10.1007/s10639-024-13263-0

Jia, X.-H., & Tu, J.-C. (2024). Towards a New Conceptual Model of AI-Enhanced Learning for College Students: The Roles of Artificial Intelligence Capabilities, General Self-Efficacy, Learning Motivation, and Critical Thinking Awareness. *Systems*, *12*(3), 74. https://doi.org/10.3390/systems12030074

Khan, A., Hughes, J., Valentine, D., Ruis, L., Sachan, K., Radhakrishnan, A., Grefenstette, E., Bowman, S. R., Rocktäschel, T., & Perez, E. (2024). *Debating with More Persuasive LLMs Leads to More Truthful Answers* (arXiv:2402.06782). arXiv. https://doi.org/10.48550/arXiv.2402.06782

Kim, J., Yang, N., & Jung, K. (2024, August 16). *Persona is a Double-edged Sword: Mitigating the Negative Impact of Role-playing Prompts in Zero-shot Reasoning Tasks*. https://www.semanticscholar.org/paper/Persona-is-a-Double-edged-Sword%3A-Mitigating-the-of-Kim-Yang/1d36a4d23c5866c9aa69f99a7f53026504c3c75a

Kong, A., Zhao, S., Chen, H., Li, Q., Qin, Y., Sun, R., Zhou, X., Wang, E., & Dong, X. (2023). *Better Zero-Shot Reasoning with Role-Play Prompting*. https://doi.org/10.48550/ARXIV.2308.07702

Kuka, V. (2024, September 27). *Role Prompting: Guide LLMs with Persona-Based Tasks*. https://learnprompting.org/docs/advanced/zero_shot/role_prompting

Lawasi, M. C., Rohman, V. A., & Shoreamanis, M. (2024). The Use of AI in Improving Student's Critical Thinking Skills. *Proceedings Series on Social Sciences & Humanities*, *18*, 366–370. https://doi.org/10.30595/pssh.v18i.1279

Morris, M. R., Sohl-dickstein, J., Fiedel, N., Warkentin, T., Dafoe, A., Faust, A., Farabet, C., & Legg, S. (2024). *Levels of AGI for Operationalizing Progress on the Path to AGI* (arXiv:2311.02462). arXiv. https://doi.org/10.48550/arXiv.2311.02462

Salter, M., & Bonfield, C. (2024). Making Human Learning Visible in a World of Invisible Generative AI: An International Perspective. *ASCILITE Publications*, 95–96. https://doi.org/10.14742/apubs.2024.1128

Stuchlikova, L., & Weis, M. (2024). From Information to Insight: Reimagining Critical Thinking Pedagogy in the Age of Artificial Intelligence. *2024 International Conference on Emerging eLearning Technologies and Applications (ICETA)*, 591–598. https://doi.org/10.1109/ICETA63795.2024.10850787

Toscano, R., Guerra, M. A., Durán-Ballén, S., & Valarezo, B. M. (2024). WIP - Development of Critical Thinking in AEC Students Aided by Artificial Intelligence. *2024 IEEE Frontiers in Education Conference (FIE)*, 1–6. https://doi.org/10.1109/FIE61694.2024.10893092

Zhou, X., Teng, D., & Al-Samarraie, H. (2024). The Mediating Role of Generative AI Self-Regulation on Students' Critical Thinking and Problem-Solving. *Education Sciences*, *14*(12), Article 12. https://doi.org/10.3390/educsci14121302

Authors

Dr Jack Tsao

Jack is the Associate Director and Senior Lecturer at the Common Core at the University of Hong Kong. With a wealth of experience in social impact, experiential learning, and project-based initiatives, Jack focuses on promoting transdisciplinarity and preparing students for the future of work and citizenship. As a Senior Fellow of Higher Education Advance, he is passionate about incorporating innovative teaching methods like gaming, storytelling, and digital and artificial intelligence technologies. He earned his PhD in Education from the University of Queensland and specialises in educational futures through research on curriculum and pedagogy, comparative and international education, artificial intelligence, and the sociology of education.

Dr Alice Sook Mun Wong

Alice is a Senior Research Assistant at the Common Core University of Hong Kong and will transition to her new role as Assistant Lecturer at the AI and Humanities Lab (University of Hong Kong). She earned her PhD in Computational and Data Sciences from Chapman University and holds a BA in Psychology with a Cognitive Science minor from UCLA. Her interdisciplinary research focuses on human agency, neuroscience, and the philosophy of Artificial Intelligence.

The Teacher Development Grant at the University of Hong Kong supported the development of this guidebook.