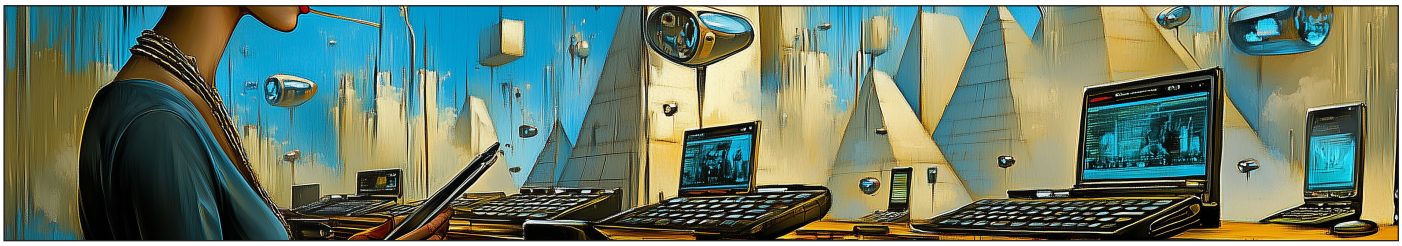
Studies in Technology Enhanced Learning

Journal homepage: **stel.pubpub.org**

Article type
Full paper, double-blind peer review.

Publication history
Received: 04 December 2024. Revised: 16 December 2024. Accepted: 16 December 2024. Online: 03 February 024.

Cover image
Badly Disguised Bligh via flickr.



Special issue Generative AI and education | **More at** https://doi.org/10.21428/8c225f6e.0390b853

Higher order prompting: Applying Bloom's revised taxonomy to the use of large language models in higher education

Jonathan Jackson

Faculty of Science and Engineering, Queen Mary University of London, London, United Kingdom

Keywords
Large Language Models; LLMs; Generative AI; GenAI; Artificial Intelligence; Prompt Engineering; Higher Education; Bloom's Taxonomy

Citation
Jackson, J. (in press). Higher order prompting: Applying Bloom's revised taxonomy to the use of large language models in higher education. **Studies in Technology Enhanced Learning, 4**(1), 1-17.

https://doi.org/10.21428/8c225f6e.0915c17e

Abstract

Large Language Models (LLMs) and Generative AI (GenAI) tools have been popularised through the widely publicised launch of tools such as ChatGPT by OpenAI and Copilot by Microsoft. While being able to generate seemingly humanlike language and respond to human input in an apparently conversational manner, these tools are not able to understand problems in the same way that humans do, and their outputs are based on probabilistic algorithms and are susceptible to hallucinations, even with extensive training. As with any tool, LLMs can be used poorly and even unethically, but there are benefits to be gained through comprehension of how LLMs work and how prompt engineering can drastically improve the outputs from LLMs. This paper builds on existing research in the field of prompt engineering while engaging with the ongoing discourse of how AI is impacting higher education and wider society. Utilising Bloom's revised taxonomy of learning as a foundation, a theoretical model is proposed for the purpose of helping educators analyse and evaluate what types of LLM usage may be appropriate within subject-specific learning experiences. Practical examples of prompts are provided as a starting point for academics and students to use and extend as part of their own experimentation with GenAI tools. The intent of this pedagogy-first approach to the use of emerging AI technologies in education

will empower students to develop their higher order thinking skills (HOTS) and help them work towards co-creation in collaboration with LLMs without compromising academic integrity or their own agency.

1. Introduction

Large Language Models (LLMs) belong to a subset of Artificial Intelligence (AI) referred to as Generative AI (GenAI) and have captured the public consciousness since ChatGPT was released by OpenAI (2022) in late 2022. LLMs such as ChatGPT have been used (and abused) in increasingly inventive ways, and the impact on Higher Education (HE) has been widely documented in the academic literature (Anson, 2024; Chiu et al., 2023; Dwivedi et al., 2023; Idris et al., 2024; Lee et al., 2024).

The application of LLMs to knowledge work such as writing, programming, and ideation, has already shown direct performance increases (Dell'Acqua et al., 2023) but in an educational context, caution is needed as heavy reliance on LLMs may hinder students from undergoing "the enculturation and cognitive development necessary for success at university" (Anson, 2024, p. 1) and, indeed, the impact could be long-lasting beyond their university studies if students' abilities to develop higher-order thinking skills (HOTS) is hindered due to a lack of appropriate guidance (Lee et al., 2024).

It is therefore critical for educators to be able to cultivate a safe learning environment that supports students' interactions with LLMs (Delcker et al., 2024) in a way that enriches their subject-specific learning experience and stimulates higher order thinking rather than eroding it. While the need to cultivate a safe learning environment is beyond dispute, there are numerous ways in which GenAI can be embedded into HE. "Insufficient knowledge of AI technologies among teachers" (Chiu et al., 2023, p. 12) may compound the issue, leading to well-meaning but ill-informed use of GenAI by educators which could subsequently be passed on to students.

Rasul et al. (2023) state that "institutions need to train students and academics in the use and misuse of ChatGPT for research and data analytics" while Chiu (2024, p. 1) calls for the transformation of HE "to train students to be future-ready for employment in a society powered by GenAI" but more discussion is still needed to ascertain "the appropriate procedures for establishing guidelines for the use of LLMs" (De Fine Licht, 2023, p. 1).

With the impending ubiquity of tools such as ChatGPT, educators must be empowered to make well-informed decisions about the use of LLMs (Mah & Groß, 2024). This paper aims to contribute constructively to the wider discourse around AI in HE and offer the foundations of a theoretical model which can be used by institutions to engage educators with the decision-making process of how – or indeed if – to use LLMs within context-specific learning environments.

For the sake of clarity and transparency, this paper avoids using the umbrella term "Artificial Intelligence" or "AI" beyond this introduction (unless contained within quotes from other sources) due to the problematic nature of the term itself which can all too easily be used to "obfuscate, alienate, and glamorize" (Tucker, 2022).

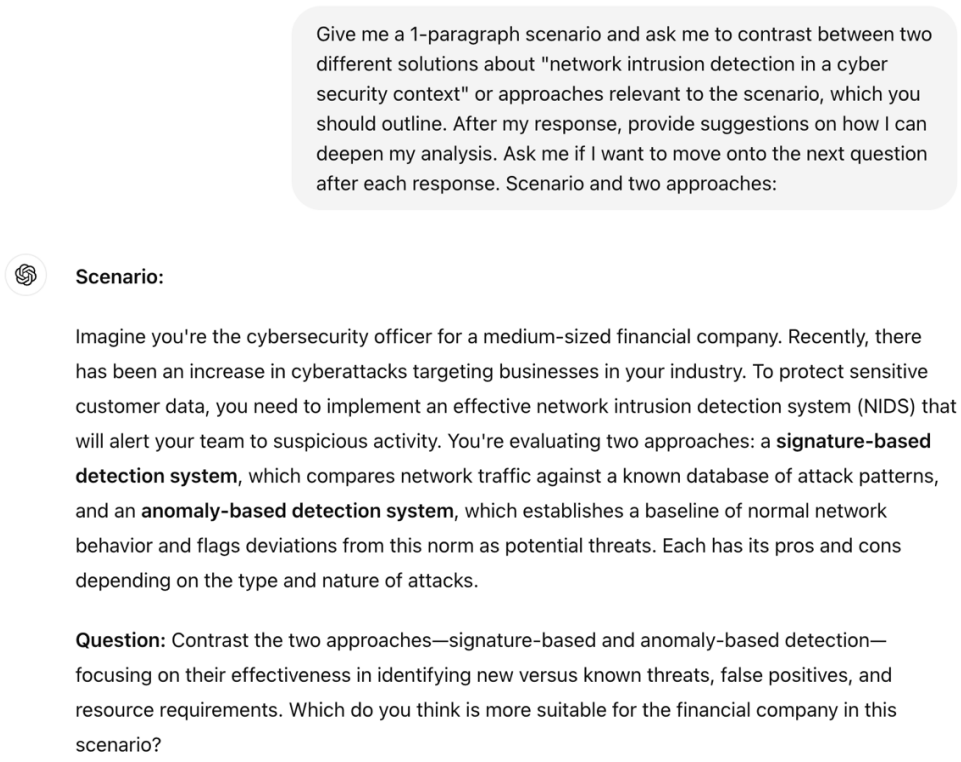
2. Large Language Models (LLMs)

LLMs are very good at generating text (a response) based on the input text (a prompt) from a user and keep track of contextual information across multiple interactions (the context window). While the results are usually impressive, LLMs can present falsities using language which conveys absolute confidence. As well as making up "facts" which are often referred to as hallucinations (just one example of anthropomorphising language which has become prevalent), LLMs can generate biased or toxic text and ignore user instructions (Ouyang et al., 2022; Wei et al., 2024). At no point are they thinking, reasoning, or understanding (Floridi, 2023; Shanahan, 2024).

Chomsky et al. (2023, p. 2) sum up LLMs adroitly: "Roughly speaking, they take huge amounts of data, search for patterns in it and become increasingly proficient at generating statistically probable outputs — such as seemingly humanlike language and thought". In a now widely cited paper, Bender et al. (2021) describe LLMs as "stochastic parrots", explaining that a LLM "is a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning" (p. 8).

Despite certain claims of LLMs being able to "reason" and "think" (OpenAI, 2024), they are – and will always be – fundamentally different to humans, regardless of how much anthropomorphising language and functionality is deployed. It is the author's position that understanding the probabilistic nature of LLM responses (Stutz et al., 2024) is key to being able to make informed decisions about their applicability to different learning contexts.

Figure 1. An example prompt which specifies the constraints on the desired response (also shown). The "Scenario and two approaches:" text at the end of the prompt is an example of a "simple colon prompt" as per Reynolds and McDonell (2021). The "flipped interaction" technique (White et al., 2023) is also evidenced through the instruction provided to the LLM to drive the conversation by asking the user questions. LLM used: ChatGPT web version, GPT-4o.



Dell'Acqua et al. (2023, p. 3) highlight that the capabilities of GenAI tools such as LLMs are expanding in an uneven manner, creating a "jagged technological frontier" which is difficult to navigate even for professionals dedicated to exploring it. The key point here is that some tasks (within the frontier) will benefit from productivity gains, but some tasks (outside of the frontier) will suffer from decreased performance. This frontier changes over time as LLMs with different capabilities emerge. This makes it even more important for educators to test GenAI tools within specific contexts before advocating their use to students, particularly because the impact will be on students' learning and cognitive developments rather than the arguably more mundane measure of worker productivity.

3. Prompt engineering

Despite the ever-changing LLM landscape, there are some fundamental principles that have remained applicable throughout. One example is that of "prompt engineering", which is a deliberate process of "designing, refining, and implementing prompts or instructions that guide the output of LLMs to help in various tasks" (Meskó, 2023, p. 1) or "programming in natural language" (Reynolds & McDonell, 2021, p. 4).

One of the foundational steps in prompt engineering is clearly defining the specific type of response that is desired (Cain, 2024). A simple example of this might be appending "Respond with a maximum of three sentences" to a prompt, which would (usually) constrain the length of the output generated by the LLM.

The way questions are worded can also be classified as prompt engineering in the sense that a simply worded question will be more likely to return a simply worded response, for example. As Cain (2024) illustrates, the prompt "Explain gravity" might result in a basic explanation from an LLM appropriate for a non-specialist audience but "Explain Einstein's theory of general relativity" may result in a more technical explanation. Even changing a single word in a prompt might result in substantial changes in output or performance (Liu et al., 2023).

https://doi.org/10.21428/8c225f6e.0915c17e

Another approach involves using prompts "containing the following components: context, general request, how the [model] is to act, and output format" (Khun, 2023, p. 3). The use of examples within a prompt can help improve the performance of LLM responses in particular use cases. Any prompt which does not contain any explicit examples of the type of responses desired is described by Brown et al. (2020) as a Zero-Shot (0S) prompt, whereas prompts which contains one or many embedded examples are described as One-Shot (1S) or Few-Shot (FS) respectively. While One-Shot and Few-Shot prompting may result in improved performance when specific response types are needed, their "performance can be matched or exceeded by simple 0-shot prompts" (Reynolds & McDonell, 2021, p. 2) where clear instructions are given to reinforce the desired behaviour of the LLM (Wang et al., 2023). Techniques can be used in combination, as shown in Figure 1, which shows an example of both "simple colon" and "flipped interaction" prompting (White et al., 2023).

Another technique that has emerged with the proliferation of LLMs is "chain-of-thought prompting", which instructs the LLM to output a series of intermediate reasoning steps rather than just an answer. This can lead to significantly more accurate responses to more complex questions (Wei et al., 2022) and a relatively trivial example of this is demonstrated in Figure 2 and in Figure 3. The performance gains are particularly pronounced in the context of scientific tasks, and chain-of-thought behaviour has been embedded as standard in models such as OpenAI o1 (Jones, 2024) to leverage these gains, but these models are still susceptible to problematic behaviour such as hallucinations due to the fundamental way LLMs work.

Many other prompting techniques have been documented in the literature emerging around LLMs (White et al., 2023) and the purpose of this paper is not to provide an exhaustive overview of these techniques. Rather, armed with the knowledge that LLMs can produce such a variety of outputs depending on the prompts they receive, educators are encouraged to ask what prompts are most likely to result in the desired outputs (Reynolds & McDonell, 2021). To discover the answer to this question in specific educational contexts, educators need to experiment, test and share their findings to contribute to pedagogical innovation within higher education which might be supported by LLMs (Pratschke, 2024, p. 113). But aimless experimentation should be minimised. The sheer number of GenAI tools and the different ways of interacting with them may seem overwhelming and could distract from what should be the fundamental concern: pedagogical benefit.

Figure 2. An example of an LLM failing to answer accurately a relatively straightforward question for humans. LLM used: ChatGPT, GPT-4o via iPhone app.

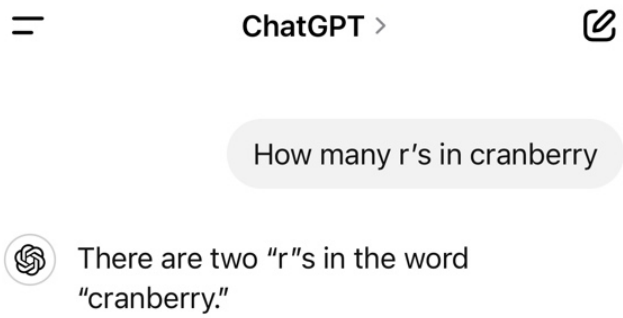
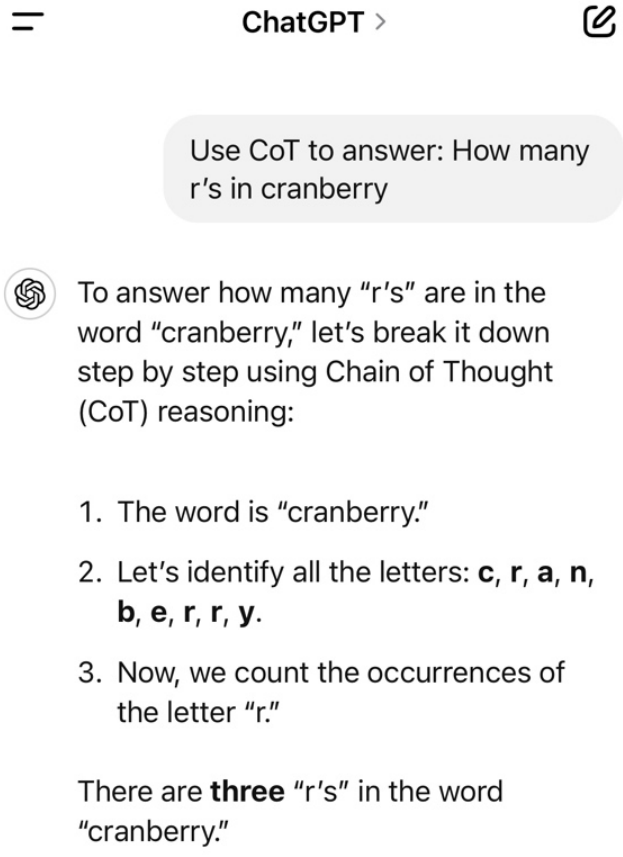


Figure 3. An example prompt requesting the LLM to utilise "chain-of-thought" which improves the accuracy of the response in certain contexts. LLM used: ChatGPT, GPT-4o via iPhone app.



So, even before considering what LLM or prompts to use, educators should reflect on what kind of learning experiences will support the development of HOTS, critical thinking, and cognitive development in students. A pedagogy-first approach is needed, instead of a technology-first approach (Hampel & Keil-Slawik, 2001).

4. Pedagogical benefit

As highlighted already, LLMs are capable of impressive feats of linguistic gymnastics and anthropomorphic behaviour, generating text that masquerades as the result of cognition. Demonstrating an LLM to a class of students is an act laden with significance and can lead to different impressions being made such as "hey, this can write my essays for me" (pedagogically harmful) or "hey, I can improve my learning with this" (pedagogically beneficial), or something else entirely.

Khun (2023, p. 7) talks of GenAI tools as "new pedagogical instruments" with the potential "to improve… lessons and increase student engagement", but these benefits can only be realised if LLM use is aligned with the core objectives of HE which includes developing students' critical thinking (De Fine Licht, 2023) and higher-order thinking skills in general which are seen as "indispensable for navigating the complexities of modern society" (Lee et al., 2024).

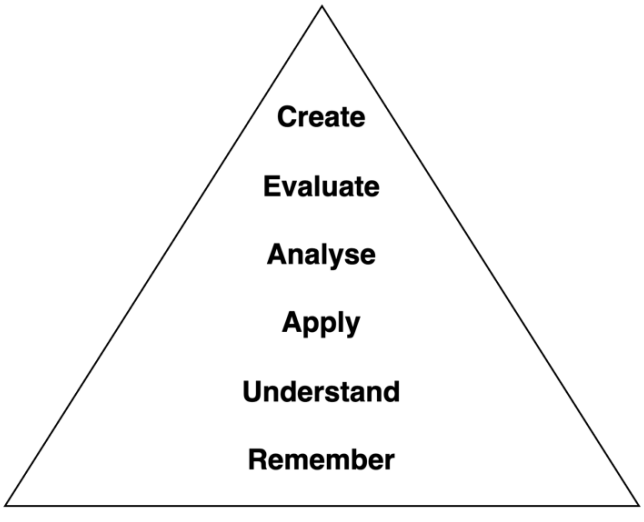
Bloom's revised taxonomy of learning for the cognitive domain, as shown in a visualisation adapted from Armstrong (2010) in Figure 4, offers a well-established and familiar "common language about learning goals" (Krathwohl, 2002, p. 1) for educators to utilise when designing learning objectives and can offer a useful lens through which to evaluate the use of LLMs in educational contexts.

In relation to Bloom's revised taxonomy, examples of higher-order thinking skills would be "evaluate" or "create" whereas "remember" or "understand" could be classified as lower-order thinking skills. Anson (2024, p. 6) argues that "LLMs have the potential to disrupt what was once considered a relatively stable hierarchy of cognitive complexity" and Rivers and Holland (2023) go so far as to suggest that LLMs invert the hierarchy altogether. They state that the "creation of knowledge is now available to the masses with the correct input of a simple prompt", but this seems to be dangerously reductive, particularly when scrutinised more carefully in contexts such as scientific knowledge production (Messeri & Crockett, 2024). The generation of text *en masse* is not the same as intentionally creating new knowledge.

The intuitively simple way of interacting with LLM tools like ChatGPT can also obscure the need for practical and theoretical training (Delcker et al., 2024), without which students may not appreciate the wider benefits of LLMs and may even form problematic mental models, leading to pedagogical harm. If a student's early interactions with an LLM was to generate essay paragraphs for an assignment, for example, they may not be aware of what else it can do. This is where educators can, and should, step in.

It is also important to note that use of LLMs within an education context is not a solely binary decision, as there are varying degrees of LLM use and reliance (Furze et al., 2024). In the context of writing, Knowles (2024, p. 1) presents Rhetorical Load Sharing "as a theoretical framework for placing texts on a collaborative authorship spectrum spanning from human-authored text to synthetic text". As part of this spectrum, "human-in-the-loop" usage is contrasted with "machine-in-the-loop" usage, and this has been incorporated into Figure 5, which visualises the different degrees of LLM usage that might exist in a wider educational context.

Figure 4. Visualisation of Bloom's revised taxonomy adapted from Armstrong (2010)



https://doi.org/10.21428/8c225f6e.0915c17e

LLM-centric usage would typically be limited to simplistic interactions involving a single prompt and a single response. Human-in-the-loop usage might be typified by using an LLM to generate the first draft of an essay and then prompting the LLM to refine it. Here, there is a chain of interaction (or conversation), and a human has been "in the loop", but the final output will be wholly synthetic. This over-reliance on LLMs can be problematic for developing higher-order thinking skills, though. A machine-in-the-loop approach counters this problem by keeping the locus of control with the user and an LLM might be used to help with specific tasks such as topic exploration, spelling, and grammar. The user remains fully in control, preserving their own intentionality (Floridi, 2023) and agency.

Taking all of these principles into consideration, it is vital to be able to identify what pedagogically beneficial use of LLMs looks like within HE, particularly in the context of Bloom's revised taxonomy. Educators also need to be supported in scaffolding appropriate use of LLMs within learning experiences in order to help students develop their HOTS.

To help address these challenges, the following initial set of pedagogical principles is proposed:

- P1. LLMs can be effective enablers of a foundational "Discovery" phase in the learning process.
- P2. Individual LLM prompts can support students across multiple levels of Bloom's revised taxonomy.
- P3. Flipped interaction prompting can increase student engagement through LLM chains of interaction but does not necessarily support HOTS.
- P4. The stochastic nature of LLMs inherently limits their reliability as sources of knowledge but can facilitate opportunities for criticality and divergent thinking.
- P5. LLM use in education should contribute to a "left shift" towards human-centric intentionality and agency as part of the development of HOTS.

The author proposes the term "Higher Order Prompting" (HOP) to encompass this set of LLM- and discipline-agnostic, pedagogy-first principles for integrating LLMs in learning contexts to support HOTS. The academic community is encouraged to revise, adapt, and add to these principles as part of context-specific scholarship.

The following sections explain these principles in more depth, along with sample LLM prompts to demonstrate how they might be applied in practice.

5. A new base for Bloom's Taxonomy

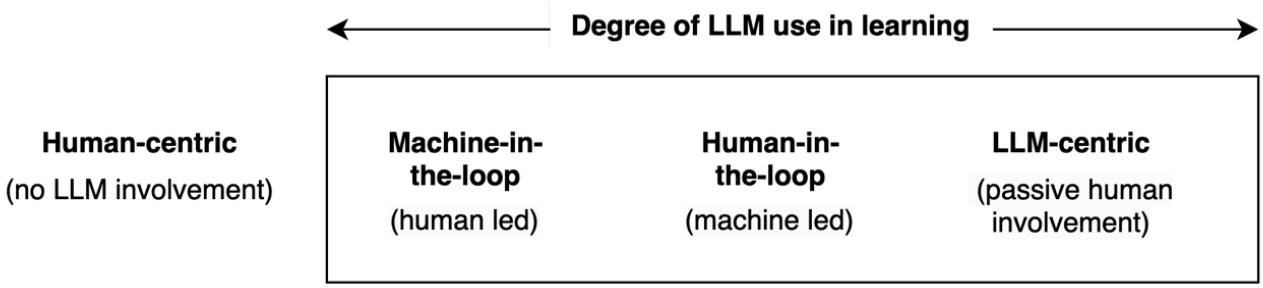
In relation to the base of Bloom's revised taxonomy (see Figure 4), there is a valid argument for not needing to remember as much as was once needed due to the availability of knowledge through online search engines (e.g., Google or DuckDuckGo) and seemingly now available through LLMs.

In fact, LLMs can be much more effective than search engines if a word cannot be recalled but its meaning can be described. For example, using the following prompt with many (but not all) LLMs would help reunite a user with the phrase "l'appel du vide":

> Prompt 1
>
> What's the term in French to do with looking over a cliff and feeling like you want to jump?

Additionally, Kasneci et al. (2023) state that LLMs can "assist in the development of research skills by providing students with information and resources on a particular topic and hinting at unexplored aspects and current research

Figure 5. The varying degrees of LLM usage and locus of control in an educational context, incorporating the concept of rhetorical load sharing from Knowles (2024).



topics". While this is certainly true, there are limitations that need to be accounted for such as information accuracy and recency. Also, student research skills will neither inevitably emerge simply by using an LLM, nor will HOTS. These skills could even be suppressed by poor LLM usage, and this is discussed further throughout this paper.

A useful starting point to support students in their understanding of topics can be as simple as asking for summaries or explanations of topics or terms, as demonstrated by Prompt 2.

> Prompt 2
>
> Explain [TOPIC] in a single sentence

Taking this further, LLMs can help students uncover the "unexplored aspects" of topics, as mentioned by Kasneci et al. (2023). In fact, LLMs are particularly good at this type of task due to the large amount of data they have typically been trained on. LLMs can subsequently "provide the basis for a very broad exploration of relevant ideas from various contexts and should be able to support divergent thinking" (Bouschery et al., 2023). Prompt 3 demonstrates a possible starting point with which to embark on this type of exploration.

> Prompt 3
>
> Tell me more about [TOPIC] and give me a bullet point list of related topics or theories which may be worth exploring

Authorial experience attests to the benefit of this approach, both for intra-topic exploration (filling gaps) and extra-topic exploration (broadening horizons). Arguably, knowledge exploration of this nature could form a new base for the adapted Armstrong (2010) visualisation of Bloom's revised taxonomy (P1) so that "Discover" might sit beneath "Remembering" in a learning environment augmented by discovery tools such as LLMs (see Figure 6).

6. Flipping the interaction paradigm

An alternative way of working with LLMs becomes apparent when the interaction paradigm is flipped by prompting LLMs to ask the user questions, as demonstrated by a very concise example in Prompt 4. This could certainly help develop a user's ability to remember facts, and may even lead to some inadvertent discovery, but in a highly ad hoc manner.

> Prompt 4
>
> Quiz me on: [TOPICS]

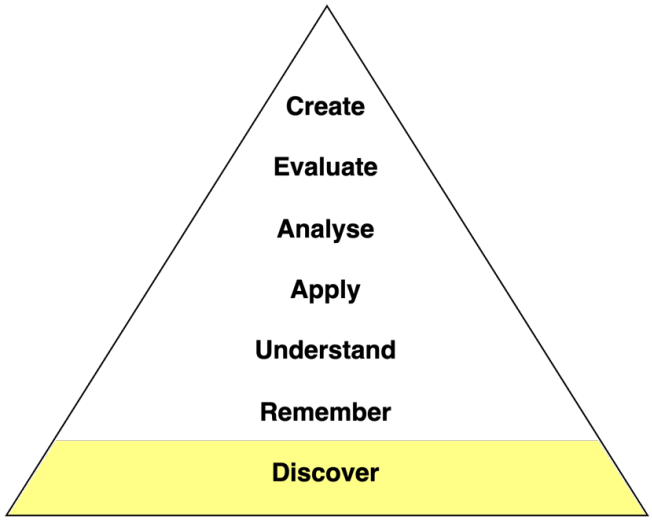
By adding more detail and behaviour re-enforcing language to the prompt, feedback can be explicitly requested from the LLM to help a user, not only with their remembering, but with their understanding as well (see Prompt 5).

> Prompt 5
>
> Quiz me using multiple choice questions on the following topic(s): [TOPICS]. Ask me one question at a time. If I get the answer wrong, give me a concise explanation about the correct answer and ask if I would like to continue with the next question. First question:

By doing so, support for both the "Remembering" and "Understanding" levels of Bloom's revised taxonomy has been combined in a single prompt, demonstrating that the levels are not isolated from each other in the context of LLM prompting (P2). This may seem self-evident, but it would seem to be an important point to make regardless, as educators look to recommend prompts (and, indeed, chains of interactions) that help students develop their HOTS.

Prompt 6 demonstrates how a prompt may be used to request mnemonic devices to aid remembering certain facts related to a topic, but of course their efficacy may vary. Not only is there an expanding number of LLMs to interact with which have different architectures and are trained on

Figure 6. Visualisation of Bloom's revised taxonomy with addition of "Discover" level at the base.



https://doi.org/10.21428/8c225f6e.0915c17e

different data, but each LLM can generate different responses to a single prompt due to the stochastic nature of LLMs. Hence, the importance of educators experimenting with LLMs to ascertain their applicability to specific contexts.

> Prompt 6
>
> Ask me to recall specific facts related to [TOPIC]. For each item, if I struggle to remember, provide hints or mnemonic devices to aid in my recall. After each correct response, briefly discuss the significance or context of the remembered information. First question:

Prompts 4, 5 and 6 have demonstrated "flipped interaction prompting" (Huber et al., 2024) which importantly reminds us that LLMs are not simply question-answering machines. In fact, using them exclusively in this manner can simultaneously increase pedagogical risk and suppress pedagogical benefits (P3).

There is inherent risk in relying on LLMs too much for knowledge recall because they are prone to hallucinations, as already highlighted. While their accuracy can be improved significantly through techniques such as Reinforcement Learning from Human Feedback (RLHF), the closer they get to 100% accuracy, the higher the risk of automation bias (Huber et al., 2024) and reduced scepticism (Knowles, 2024) in contexts where scepticism and criticality is still very much required.

By extension, flipped interaction prompting also has risks. To illustrate the point, Pham et al. (2024) reported that expert verification of mathematics content generated by ChatGPT, which included question-answer pairs, led to an approval rate of 87.70%. In essence, more than 1 in every 10 questions were deemed either incorrect or inappropriate. The authors also reported on the struggles which ChatGPT had in enhancing questions' difficulty levels (Pham et al., 2024, p. 7). Another study reported on LLM-generated questions (Elkins et al., 2023), but with a significant amount of initial effort expended in providing human-written example questions as part of a 5-shot prompting strategy. This is not a tenable approach for students in a self-directed study context. Authorial experience of running Generative AI workshops with students has also uncovered examples of LLMs generating multiple choice questions with incorrect answers in domains such as music history and sports therapy.

Of course, newer LLMs have subsequently been released since the above studies were conducted, and by the time this article is published, even more LLMs will have been released, each with their own set of enhancements that will have been measured in a variety of (sometimes seemingly arbitrary) ways. While accuracy may seem to be trending towards 100%, the fact that 100% accuracy is not something that can be realistically guaranteed means LLMs cannot be relied upon in isolation as question-answering (and even question-generating) machines.

One benefit of the flipped interaction prompting approach, though, is that it necessitates a certain level of active participation by the user. This at least gives a starting point in thinking about LLMs fostering HOTS in students. Active participation is a key ingredient but, of course, not sufficient in and of itself.

7. LLMs in support of knowledge application and analysis

By asking the question "who do we want to do the thinking here?", educators can be more intentional with their integration of LLMs into specific learning experiences and help students climb the levels of learning and develop their HOTS. Educators can continue to utilise flipped interaction prompting to scaffold tasks to help develop the skills of application and analysis, in line with the third and fourth levels of Bloom's revised taxonomy (apply and analyse).

Prompts 7 and 8 demonstrate how an LLM might be prompted to generate some information relating to a scenario or case study and then ask the user to answer a question based on this information with the goal being *knowledge application*.

> Prompt 7
>
> Ask me a question about [TOPIC] and give me some related pieces of information which I should then use in calculating or formulating an answer. Explain the correct answer if I get the answer wrong. Ask me if I want to move onto the next question after each response.

As always, results will vary, so educators need to assess the usefulness of different prompts by experimenting within their subject domains. If certain prompts perform well, there may be a case for sharing ready-made prompts with students to use as part of self-study, but it would be advantageous to support student engagement with LLMs within a synchronous learning environment, particularly in the early stages of student LLM engagement, to mitigate pedagogical risks.

> Prompt 8
>
> Generate a 2-paragraph case study on [THEME] and ask me one single sentence question at a time about how I can apply the topic(s) [TOPICS] to this case study. Provide suggestions on how I could apply my knowledge better to the case study. Ask me if I want to move onto the next question after each response. Case study and first question:

Allowing the LLM to generate a case study "on the fly" in this case leverages the generative nature of LLMs and it may suggest novel situations beneficial to learning which would not have arisen through manual case study creation. That said, more control can be exerted by educators by providing a fixed scenario or case study in the original prompt so that the LLM will generate questions based on this specific example. A hybrid approach may be adopted by educators if they utilise an LLM to generate a series of case studies and then review, refine, or discard them prior to sharing with students.

A prompt which might be designed to help students exercise their *analytical skills* could utilise specific language to constrain the type of analysis that is expected. For example, Prompt 9 uses the specific term "contrast" as a type of analysis. (Refer back to Figure 1 for an example response by ChatGPT to this prompt.)

> Prompt 9
>
> Give me a 1-paragraph scenario and ask me to contrast between two different solutions about [TOPIC] or approaches relevant to the scenario, which you should outline. After my response, provide suggestions on how I can deepen my analysis. Ask me if I want to move onto the next question after each response. Scenario and two approaches:

Users can, of course, ask the LLM to do the contrasting, the summarising, the calculating, the "thinking". There are obvious risks that this type of LLM-centric approach which might easily turn the student into a passive observer as the LLM performs linguistic acrobatics, but there can be benefits, particularly if the student requests multiple responses to the same prompt and then evaluates those critically.

Another potential use case is the use of LLMs to simulate stakeholder interaction in a learning environment. LLMs can be instructed to mimic certain behavioural traits by utilising "task specification by memetic proxy", as described by Reynolds and McDonell (2021). For example, including instructions such as "behave like a difficult client" can lead to teachable moments in the context of communication and negotiation (Jackson et al., 2024). Multiple LLM chains of interaction (or "chats") can be initiated with different persona information so that students can interact with different personas in a simulated project context. There is, of course, the risk of stereotypical response generation due to biases present in the LLM training data (Brown et al., 2020), but this itself should afford an opportunity for evaluative criticality with appropriate educator guidance (P4).

8. LLMs in support of critical evaluation

In the context of the fifth level of Armstrong's (2010) visualisation of Bloom's revised taxonomy, can an LLM evaluate or pass judgement on a complex scenario? An LLM-generated response may imply that the LLM is more than capable of doing so, but this highlights an area of high pedagogical risk where an important question to ask again is, "who do we want to do the thinking here?".

Stutz et al. (2024, p. 7) contend that institutions particularly "need to address the responsible use of AI and… equip students with the ability to evaluate the plausibility and veracity of AI results". Again, the educator's role is key in scaffolding appropriate activities that support students' development of HOTS, of which evaluation plays a significant part.

An example of this may be to use an LLM to generate example argumentation on a particular topic from different perspectives (see Prompt 10) and then ask students to evaluate the effectiveness of each argument, make recommendations on how they can be improved, whether they are factually reliable, and reflect on whether the LLM-generated arguments have any impact on their own perspective. This kind of task can enable educators to demonstrate the strengths and limitations of LLMs, as well as exposing students to diverse argumentation, potentially helping them to appreciate the benefits of divergent thinking (Bouschery et al., 2023).

> Prompt 10
>
> Generate three evaluative arguments taking different stances on the topic of [TOPIC]. Each argument should be two paragraphs long and be backed up by evidence where possible.

https://doi.org/10.21428/8c225f6e.0915c17e

Depending on the LLM being used, appropriate references may or may not be provided in the LLM's response. While earlier versions of ChatGPT, for example, might hallucinate (fabricate) references, as of late 2024, the freely available version of ChatGPT can reference online sources if explicitly asked (e.g., by appending "Cite and link to online sources" to the prompt). Students should be taught to question the quality and applicability of these sources as well as the lack of transparency around why particular sources are selected as part of an LLM response (Shah & Bender, 2024).

In terms of the locus of control (see Figure 5), while the generation of the original arguments could be viewed as LLM-centric, the intent of the overall activity results in a "left shift" toward the human-centric end of the scale and offers a significant opportunity for the development of HOTS through an evaluative task which could be carried out offline without the aid of an LLM (P5).

It should also not be assumed that an LLM will interpret what is meant by evaluation correctly, or even differentiate appropriately between analysis and evaluation, so this can be made explicit, as demonstrated in Prompt 11, which links evaluation to taking a position on a question and adopts a flipped interaction approach.

> Prompt 11
>
> Give me a 3-paragraph case study about [TOPIC] and ask me to present an argument backed up by evidence regarding my position on a question related to the case study. Wait for my response. Only if I ask for help, provide suggestions on how I can deepen my evaluation after I respond and suggest topics or approaches to help me refine my answer. Case study:

9. (Co-)creation with LLMs

In the context of creating novel outputs, there is an argument to be made that LLMs can support students in producing "more elaborate outcomes in a shorter time" (Stutz et al., 2024, p. 7), but saving time should not be framed as the primary motivator for using LLMs as this could easily be misinterpreted as meaning less time is needed for thinking, which could undermine the efforts of developing HOTS. Used appropriately, LLMs can provide *more* for a student to think about. Any framing of LLMs as offering a "shortcut" in learning is problematic.

As is implied by the visualisation of Bloom's revised taxonomy (see Figure 6), creation is supported by all levels beneath it, and a mix of suitable LLM prompts and chains of interaction can help students refine their own ideas as they work towards creative outputs.

More detailed prompting, which includes clear task specification and contextual information, can lead to an effective machine-in-the-loop approach to co-creation, as demonstrated in Prompt 12 which might be an effective starting point for a longer chain of interaction between the student and LLM.

> Prompt 12
>
> I am working on a report to do with [TOPIC] and I would like support uncovering topics that may be relevant which I have not included so far. First, ask me for a summary of my current report, then make suggestions and continue asking questions about areas I could improve. Ask me if I want help exploring any of these topics in more detail and then provide a concise summary of the topic(s) and suggest what related topics or academic sources I could investigate.

As alluded to previously, educators should aim to support students to "left shift" their activities and thinking toward the human-centric end of the LLM usage scale, as visualised in Figure 5, particularly for tasks which necessitate HOTS beyond the lower levels of Bloom's revised taxonomy.

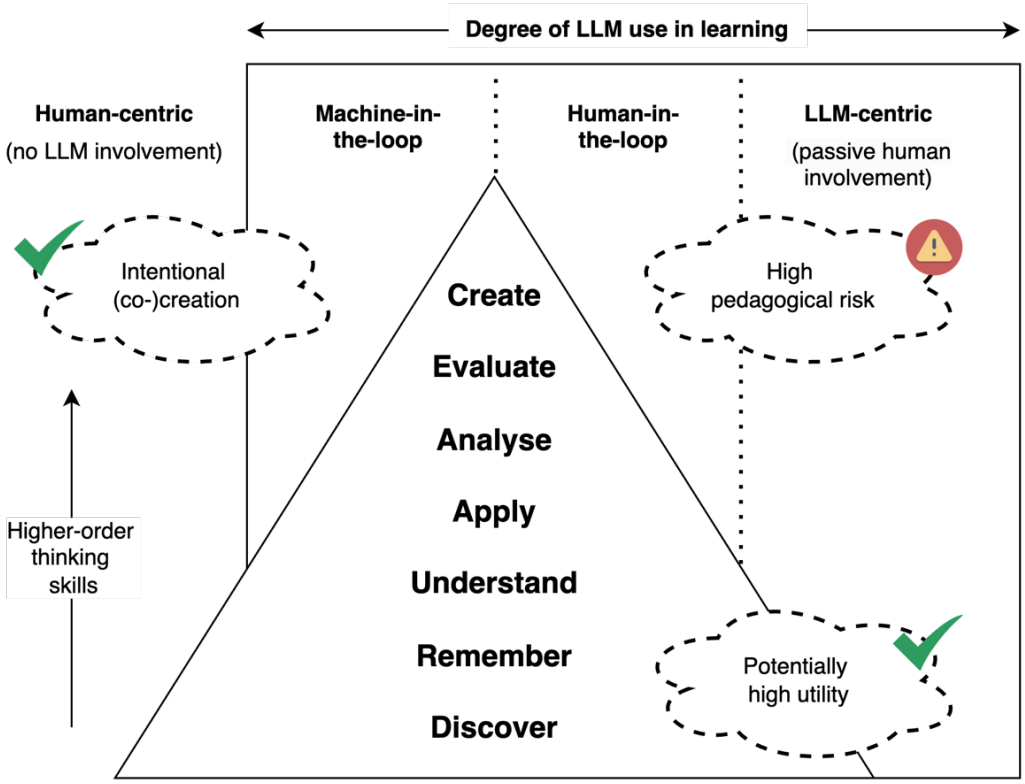
10. Higher Order Prompting

Fundamentally, HOP promotes the use of LLMs as amplifiers of human agency (Idris et al., 2024) and never as a replacement for intentional creation, where part of the creative act includes the decision regarding whether something should even be created in the first place (Floridi, 2023).

Figure 7 presents the HOP model, a combined visualisation of the key concepts covered in this paper, highlighting the potential areas of high utility and high risk when using LLMs within an educational context. The dotted lines used in the diagram indicate boundaries which are fluid and might vary depending on context-specific factors and perspectives.

While LLM-centric prompting may be a suitable starting point anywhere from discovery through to asking an LLM to evaluate or even create sample content, the intent should invariably be to shift left towards human-centric activity. In

Figure 7. Higher Order Prompting model which combines the adapted version of Bloom's revised taxonomy proposed in this paper with aspects of Knowles's (2024) rhetorical load sharing.



practice, this means that if content has been created using an LLM, there should be an active human-centric evaluation carried out as part of the process in order to support the development of HOTS.

In relation to HOTS such as evaluation and creative acts, there is a high risk of pedagogical harm if LLM use does not shift left and escape the LLM-centric or human-in-the-loop end of the scale. While LLMs can help students "expand their knowledge, capabilities, and ideas", LLMs should not be used to solve or create things on their behalf (Meskó, 2023, p. 5) as this would typically shortcut learning and be like using a car to improve physical fitness. LLMs can instead be viewed like a bicycle which still requires exertion but can take the rider further than if they walked (Bedington et al., 2024). But, of course, this does not mean a person should ride a bicycle everywhere they go. Similarly, human-centric activities with no LLM involvement are wholly appropriate in many circumstances.

It is proposed that the HOP model and the associated examples presented in this paper can be used to help educators reflect on where their own LLM prompts might sit in terms of supporting the development of HOTS in students. For example, if an educator is helping students use an LLM to discover topics related to a central theme, how might a learning activity be designed to help the students "shift left" and use the results of the LLM-centric discovery phase to move towards a human-centric understanding and application of the topics? This may involve a series of intermediary activities that sit at various points on the LLM usage scale, but the overall direction of travel should be towards the human-centric end of the scale, which is where HOTS are built.

The HOP model can also help educators reflect on whether their scaffolding of LLM-supported activities can be more diverse in their focus. Rather than relying only on LLM-centric prompting, could a variety of prompting techniques (e.g., flipped interaction prompting) be used to support a "shift left" direction of travel through human-in-the-loop and machine-in-the-loop prompting towards the human-centric building of HOTS?

The underlying question in all cases should be whether we as educators are designing activities which stimulate or hinder HOTS.

11. Ethical considerations

Algorithmic bias, the digital divide, health and safety concerns (Stahl et al., 2023), environmental and financial costs (Bender et al., 2021), toxic language generation (Gehman et al., 2020), and stereotypical or prejudicial content (Brown et al., 2020) are just some of the ethical factors and risks to consider when using LLMs, regardless of context (Weidinger et al., 2021).

The ethical factors related to GenAI technologies such as LLMs are often included as a fundamental aspect of "AI Literacy" (Ng et al., 2021), and rightly so. An important part of this literacy is to help students make informed decisions, not just about how to use GenAI, but whether to use it at all. If educators are truly intent on helping students develop HOTS, student agency and freedom of choice should also be a priority. It would seem reasonable to remind students that they can opt out of using GenAI tools if they wish, in line with valid ethical concerns. But how might this, in turn, impact their learning if LLM-supported activities are included as core components of a learning experience? Should educators ensure that alternative activities are always available? Should LLM-supported activities ever be mandatory? These would seem to be open questions worthy of future research.

12. Limitations

As a conceptual piece, this paper has the limitation of not being directly supported by any primary data, but the next steps outlined in the next section aim to build upon the HOP principles by showing how they can be applied in a live learning environment. It is also acknowledged that the LLM interaction examples presented were limited to ChatGPT only (Figures 1, 2 and 3) whereas other LLM tools such as Microsoft Copilot, Claude, Gemini, or Perplexity would be just as relevant. A comparative analysis of different LLMs was not a goal of this paper, but rather the concepts and principles presented are designed to be LLM-agnostic.

13. Next Steps

Next steps will include the design and evaluation of LLM-enhanced learning activities within an undergraduate degree module delivered at the author's institution in line with appropriate AI literacy guidelines (Zhou & Schofield, 2024). Evaluative measures are expected to include qualitative student perspectives as a minimum.

14. Conclusion

The proliferation and impending ubiquity of LLM tools such as ChatGPT has given rise to a significant discourse on the impacts of LLMs on HE, and education in general. While there are potential benefits to learning, educators must be empowered to make well-informed decisions about the use of LLMs within context-specific learning environments, balancing various ethical and pedagogical factors.

The capabilities of GenAI tools such as LLMs tend to expand in an uneven manner, creating a "jagged technological frontier", which can be difficult to track and therefore specific use cases for LLMs require experimentation underpinned by a foundational knowledge of the probabilistic nature of LLMs and their inherent limitations. They are unable to "reason" or "think" as humans do, and anthropomorphising language which implies the contrary is potentially problematic.

This paper presents a Higher Order Prompting model and set of related principles in order to help educators reflect on their own practice and how LLMs can help their students develop higher-order thinking skills. Use of this model is underpinned by approaches such as "prompt engineering" as a deliberate process of optimising LLM outputs.

As part of the Higher Order Prompting model, Bloom's revised taxonomy of learning for the cognitive domain is modified to include a foundational 'discovery' level and is recommended as a useful lens through which to evaluate the use of LLMs in educational contexts. The model can also help educators reflect on whether LLM-supported activities facilitate a "shift left" directionality towards human-centric higher-order thinking.

Within the context of this paper, the underlying question for educators is "who do we want to do the thinking here?" and whether educators are designing LLM-supported activities which stimulate or hinder higher-order thinking skills.

References

Anson, D. W. J. (2024). The impact of large language models on university students' literacy development: A dialogue with Lea and Street's academic literacies framework. *Higher Education Research & Development*, 1–14. https://doi.org/10.1080/07294360.2024.2332259

Armstrong, P. (2010). *Bloom's Taxonomy*. Vanderbilt University Center for Teaching. https://cft.vanderbilt.edu/guides-sub-pages/blooms-taxonomy/

Bedington, A., Halcomb, E. F., McKee, H. A., Sargent, T., & Smith, A. (2024). Writing with generative AI and human-machine teaming: Insights and recommendations from faculty and students. *Computers and Composition*, *71*, 102833. https://doi.org/10.1016/j.compcom.2024.102833

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜. *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://doi.org/10.1145/3442188.3445922

Bouschery, S. G., Blazevic, V., & Piller, F. T. (2023). Augmenting human innovation teams with artificial intelligence: Exploring transformer-based language models. *Journal of Product Innovation Management*, *40*(2), 139–153. https://doi.org/10.1111/jpim.12656

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., … Amodei, D. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, *33*, 1877–1901. https://proceedings.neurips.cc/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html

Cain, W. (2024). Prompting Change: Exploring Prompt Engineering in Large Language Model AI and Its Potential to Transform Education. *TechTrends*, *68*(1), 47–57. https://doi.org/10.1007/s11528-023-00896-0

Chiu, T. K. F. (2024). Future research recommendations for transforming higher education with generative AI. *Computers and Education: Artificial Intelligence*, *6*, 100197. https://doi.org/10.1016/j.caeai.2023.100197

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). Opinion | Noam Chomsky: The False Promise of ChatGPT. *The New York Times*. https://www.nytimes.com/2023/03/08/opinion/noam-chomsky-chatgpt-ai.html

De Fine Licht, K. (2023). *Integrating Large Language Models into Higher Education: Guidelines for Effective Implementation*. 65. https://doi.org/10.3390/cmsf2023008065

Delcker, J., Heil, J., Ifenthaler, D., Seufert, S., & Spirgi, L. (2024). First-year students AI-competence as a predictor for intended and de facto use of AI-tools for supporting learning processes in higher education. *International Journal of Educational Technology in Higher Education*, *21*(1), 18. https://doi.org/10.1186/s41239-024-00452-7

Dell'Acqua, F., McFowland III, E., Mollick, E. R., Lifshitz-Assaf, H., Kellogg, K., Rajendran, S., Krayer, L., Candelon, F., & Lakhani, K. R. (2023). *Navigating the Jagged Technological Frontier: Field Experimental Evidence of the Effects of AI on Knowledge Worker Productivity and Quality* (SSRN Scholarly Paper 4573321). https://doi.org/10.2139/ssrn.4573321

Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, A. S., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., … Wright, R. (2023). Opinion Paper: "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, *71*, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642

Elkins, S., Kochmar, E., Serban, I., & Cheung, J. C. K. (2023). *How Useful Are Educational Questions Generated by Large Language Models?* (N. Wang, G. Rebolledo-Mendez, V. Dimitrova, N. Matsuda, & O. C. Santos, Eds.; pp. 536–542). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-36336-8_83

Floridi, L. (2023). AI as Agency Without Intelligence: On ChatGPT, Large Language Models, and Other Generative Models. *Philosophy & Technology*, *36*(1), 15. https://doi.org/10.1007/s13347-023-00621-y

Furze, L., Perkins, M., Roe, J., & MacVaugh, J. (2024). The AI Assessment Scale (AIAS) in action: A pilot implementation of GenAI-supported assessment. *Australasian Journal of Educational Technology*, *40*(4), 38–55. https://doi.org/10.14742/ajet.9434

Gehman, S., Gururangan, S., Sap, M., Choi, Y., & Smith, N. A. (2020). RealToxicityPrompts: Evaluating Neural Toxic Degeneration in Language Models. *Findings of the Association for Computational Linguistics: EMNLP 2020*, 3356–3369. https://doi.org/10.18653/v1/2020.findings-emnlp.301

Hampel, T., & Keil-Slawik, R. (2001). sTeam: Structuring information in team-distributed knowledge management in cooperative learning environments. *Journal on Educational Resources in Computing*, *1*(2es), 3-es. https://doi.org/10.1145/384055.384058

Huber, S. E., Kiili, K., Nebel, S., Ryan, R. M., Sailer, M., & Ninaus, M. (2024). Leveraging the Potential of Large Language Models in Education Through Playful and Game-Based Learning. *Educational Psychology Review*, *36*(1), 25. https://doi.org/10.1007/s10648-024-09868-z

Idris, M. D., Feng, X., & Dyo, V. (2024). Revolutionizing Higher Education: Unleashing the Potential of Large Language Models for Strategic Transformation. *IEEE Access*, *12*, 67738–67757. https://doi.org/10.1109/ACCESS.2024.3400164

Jackson, J., Day, N., & Maher, K. (2024). Bringing the Real World Into the Classroom Through Team-Based Live Brief Projects. *Approaches to Work-Based Learning in Higher Education: Improving Graduate Employability*, 81.

Jones, N. (2024). 'In awe': Scientists impressed by latest ChatGPT model o1. *Nature*. https://doi.org/10.1038/d41586-024-03169-9

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khun, C. (2023). Prompt Engineering in Medical Education. *International Medical Education*, *2*. https://doi.org/10.3390/ime2030019

Knowles, A. M. (2024). Machine-in-the-loop writing: Optimizing the rhetorical load. *Computers and Composition*, *71*, 102826. https://doi.org/10.1016/j.compcom.2024.102826

Krathwohl, D. R. (2002). A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice*. https://doi.org/10.1207/s15430421tip4104_2

Lee, D., Arnold, M., Srivastava, A., Plastow, K., Strelan, P., Ploeckl, F., Lekkas, D., & Palmer, E. (2024). The impact of generative AI on higher education learning and teaching: A study of educators' perspectives. *Computers and Education: Artificial Intelligence*, *6*, 100221. https://doi.org/10.1016/j.caeai.2024.100221

Liu, X., Zheng, Y., Du, Z., Ding, M., Qian, Y., Yang, Z., & Tang, J. (2023). GPT understands, too. *AI Open*. https://doi.org/10.1016/j.aiopen.2023.08.012

Mah, D.-K., & Groß, N. (2024). Artificial intelligence in higher education: Exploring faculty use, self-efficacy, distinct profiles, and professional development needs. *International Journal of Educational Technology in Higher Education*, *21*(1), 58. https://doi.org/10.1186/s41239-024-00490-1

Meskó, B. (2023). Prompt Engineering as an Important Emerging Skill for Medical Professionals: Tutorial. *Journal of Medical Internet Research*, *25*(1), e50638. https://doi.org/10.2196/50638

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. *Nature*, *627*(8002), 49–58. https://doi.org/10.1038/s41586-024-07146-0

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, *2*, 100041. https://doi.org/10.1016/j.caeai.2021.100041

OpenAI. (2022, November 30). Introducing ChatGPT. *OpenAI*. https://openai.com/index/chatgpt/

OpenAI. (2024, September 12). Introducing OpenAI o1. *OpenAI*. https://openai.com/index/introducing-openai-o1-preview/

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). *Training language models to follow instructions with human feedback* (arXiv:2203.02155). arXiv. https://doi.org/10.48550/arXiv.2203.02155

Pham, P. V. L., Duc, A. V., Hoang, N. M., Do, X. L., & Luu, A. T. (2024). ChatGPT as a Math Questioner? Evaluating ChatGPT on Generating Pre-university Math Questions. *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*, 65–73. https://doi.org/10.1145/3605098.3636030

Pratschke, B. M. (2024). *Generative AI and Education: Digital Pedagogies, Teaching Innovation and Learning Design*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-67991-9

Rasul, T., Nair, S., Kalendra, D., Robin, M., de Oliveira Santini, F., Ladeira, W. J., Sun, M., Day, I., Rather, R. A., & Heathcote, L. (2023). The role of ChatGPT in higher education: Benefits, challenges, and future research directions. *Journal of Applied Learning and Teaching*, *6*(1), 41–56.

Reynolds, L., & McDonell, K. (2021). Prompt Programming for Large Language Models: Beyond the Few-Shot Paradigm. *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–7. https://doi.org/10.1145/3411763.3451760

Rivers, C., & Holland, A. (2023, August 30). How can generative AI intersect with Bloom's taxonomy? *Times Higher Education, Inside Higher Ed.* https://www.timeshighereducation.com/campus/how-can-generative-ai-intersect-blooms-taxonomy

Shah, C., & Bender, E. M. (2024). Envisioning Information Access Systems: What Makes for Good Tools and a Healthy Web? *ACM Trans. Web*, *18*(3), 33:1-33:24. https://doi.org/10.1145/3649468

Shanahan, M. (2024). Talking about Large Language Models. *Commun. ACM*, *67*(2), 68–79. https://doi.org/10.1145/3624724

Stahl, B. C., Schroeder, D., & Rodrigues, R. (2023). *Ethics of Artificial Intelligence: Case Studies and Options for Addressing Ethical Challenges*. Springer Nature. https://doi.org/10.1007/978-3-031-17040-9

Stutz, P., Elixhauser, M., Grubinger-Preiner, J., Linner, V., Reibersdorfer-Adelsberger, E., Traun, C., Wallentin, G., Wöhs, K., & Zuberbühler, T. (2024). *Ch(e)atGPT? An Anecdotal Approach addressing the Impact of ChatGPT on Teaching and Learning GIScience*. https://doi.org/10.35542/osf.io/j3m9b

Tucker, E. (2022, March 17). Artifice and Intelligence. *Tech Policy Press*. https://techpolicy.press/artifice-and-intelligence

Wang, J., Shi, E., Yu, S., Wu, Z., Ma, C., Dai, H., Yang, Q., Kang, Y., Wu, J., Hu, H., Yue, C., Zhang, H., Liu, Y., Li, X., Ge, B., Zhu, D., Yuan, Y., Shen, D., Liu, T., & Zhang, S. (2023). *Prompt Engineering for Healthcare: Methodologies and Applications* (arXiv:2304.14670). arXiv. http://arxiv.org/abs/2304.14670

Wei, J., Karina, N., Chung, H. W., Jiao, Y. J., Papay, S., Glaese, A., Schulman, J., & Fedus, W. (2024). *Measuring short-form factuality in large language models* (arXiv:2411.04368). arXiv. https://doi.org/10.48550/arXiv.2411.04368

Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q. V., & Zhou, D. (2022). Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. *Advances in Neural Information Processing Systems*, *35*, 24824–24837.

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., Kasirzadeh, A., Kenton, Z., Brown, S., Hawkins, W., Stepleton, T., Biles, C., Birhane, A., Haas, J., Rimell, L., Hendricks, L. A., … Gabriel, I. (2021). *Ethical and social risks of harm from Language Models* (arXiv:2112.04359). https://doi.org/10.48550/arXiv.2112.04359

White, J., Fu, Q., Hays, S., Sandborn, M., Olea, C., Gilbert, H., Elnashar, A., Spencer-Smith, J., & Schmidt, D. C. (2023). *A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT* (arXiv:2302.11382). https://doi.org/10.48550/arXiv.2302.11382

Zhou, X., & Schofield, L. (2024). Developing a conceptual framework for Artificial Intelligence (AI) literacy in higher education. *Journal of Learning Development in Higher Education*, *31*, Article 31. https://doi.org/10.47408/jldhe.vi31.1354



About the author

Jonathan Jackson is a Senior Lecturer in Software Engineering and Management with a background in industry and has Chartered IT Professional status from the British Computer Society (BCS). He is currently Programme Director for the Digital and Technology Solutions undergraduate and postgraduate degree apprenticeship programmes at Queen Mary University of London. Jonathan's scholarly interests include digitally enhanced teaching and learning, Large Language Models (LLMs) in education, and authentic assessment through industry collaboration.

jonathan.jackson@qmul.ac.uk **0009-0005-3437-8081**

bsky.app/profile/iamjonjackson.bsky.social

Open Access (CC BY 4.0)

© 2020 The Authors. This article is distributed under Creative Commons Attribution 4.0 International licence.

You are free to

- Share — copy and redistribute the material in any medium or format
- Adapt — remix, transform, and build upon the material for any purpose, even commercially.

Under the following terms:

- Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.
- No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

The full licence conditions are available at: **https://creativecommons.org/licenses/by/4.0/**