# What skills can students learn from using generative AI in assessments?

**By Matt Lucas and Isabel Fischer (Warwick Business School)**

*Matt Lucas is a Senior Product Manager at IBM, and Isabel Fischer Reader (Associate Professor) in Information Systems at WBS (Warwick Business School). Isabel also co-convenes an IATL module. This blog represents their own opinions and not those of their employers.*

It is rare for a technology to capture the public’s imagination in the way that AI has. With the public introduction of ChatGPT v3.5 in November 2022, people have been waking up to the potential that the Large Language Models (LLMs) that underpin many AI solutions might have on the future relationship we have with machines. After decades of research, and promises of an AI revolution “just over the horizon”, ChatGPT has kickstarted disruption of a wide number of professions in just a few short months, from traditional creatives such as musicians, authors and artists to recruiters, politicians and even doctors. [references]

Many have seen the rapid adoption and innovation of LLMs and called for tighter regulation, a pause, or even an outright ban. The [UK](https://eandt.theiet.org/content/articles/2023/05/sunak-vows-to-establish-ai-guardrails/) and [EU](https://apnews.com/article/tech-ai-artificial-intelligence-europe-eu-15ac394679519084478e15217c156abc) have independently announced plans to implement guardrails around the use of AI. In March 2023 numerous tech industry leaders called for a [pause to AI research](https://futureoflife.org/open-letter/pause-giant-ai-experiments/). In the same month, Italy [banned ChatGPT](https://www.theguardian.com/technology/2023/mar/31/italy-privacy-watchdog-bans-chatgpt-over-data-breach-concerns) completely, only to [open it up](https://www.bbc.com/news/technology-65431914) again since. In May 2023 Sam Altman, CEO of OpenAI spoke to [congress about AI regulation](https://www.bbc.co.uk/news/world-us-canada-65616866), an idea that is also suggested by Yuval Noah Harari’s talk on [AI and the future of humanity](https://www.youtube.com/watch?v=LWiM-LuRe6w), especially worrying if seen in conjunction with [‘deep fakes’](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7406872/) and also in conjunction with humans’ tendencies to sometimes [anthropomorphise AI's functionalities and innovations](https://pubmed.ncbi.nlm.nih.gov/32228388/).

Cognizant that LLMs have much wider ethical concerns, what, if any, are the positives of LLMs for lecturers? As lecturers, how can we make good use of this technology when building courses, in a way that adequately prepares students for the world in which they will work? In this blog we are going to cover some ways in which LLMs can be used in Higher Education by teachers who are looking to embrace it.

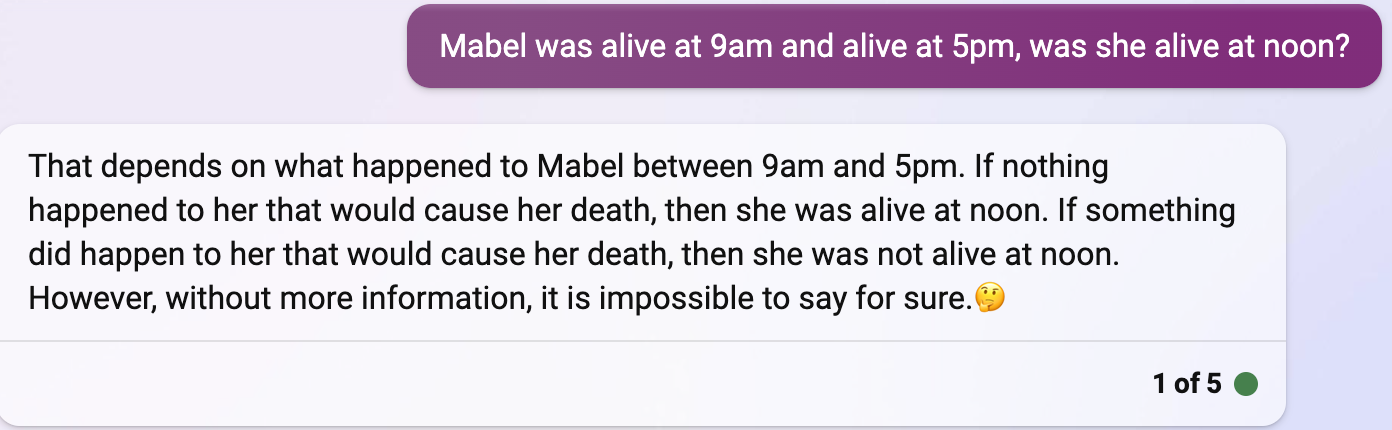
### Strengths and weaknesses of LLMs

[A recent HBR article](https://hbsp.harvard.edu/inspiring-minds/the-10-cognitive-skills-your-students-future-employers-seek?cid=email%7Cmarketo%7C2023-05-18-inspiring-minds-digest-theme-education-strategy%7C1367442%7Cthought-lead-inspiring-minds%7Ceducator%7Cinspiring-minds-article%7Cmay2023&acctID=17096278&mkt_tok=ODU1LUFUWi0yOTQAAAGLzX-IABLgdgYYQqJs2_Ueyj8wqJci2cGZMECMlz7sDTadUP00KeiYOY9s_3nzEc03rFmQbrr-an0_st0FUKLK3damVoYJfGNiYEHUN27hCA) summarises ten types of thinking skills students need: Analytical thinking, Synthetic thinking, Convergent thinking, Divergent thinking, Constructive thinking, Deconstructive thinking, Inductive thinking, Deductive thinking, Abductive thinking and Representational thinking. Our job as educators is to teach students the skills that prepare them for a world in which LLMs are commonplace, just as we accepted the use of electronic calculators in the classroom once maths students grasped the basics of mental arithmetic.

One way to approach LLMs might be to consider the skills that students can learn that complement those of LLMs. There is a lot of mainstream literature on the strengths and weaknesses of LLMs. For example, cognitive psychologist, Steven Pinker, [talks about two key weaknesses](https://www.bbc.co.uk/sounds/play/m001lyx3) of LLMs:

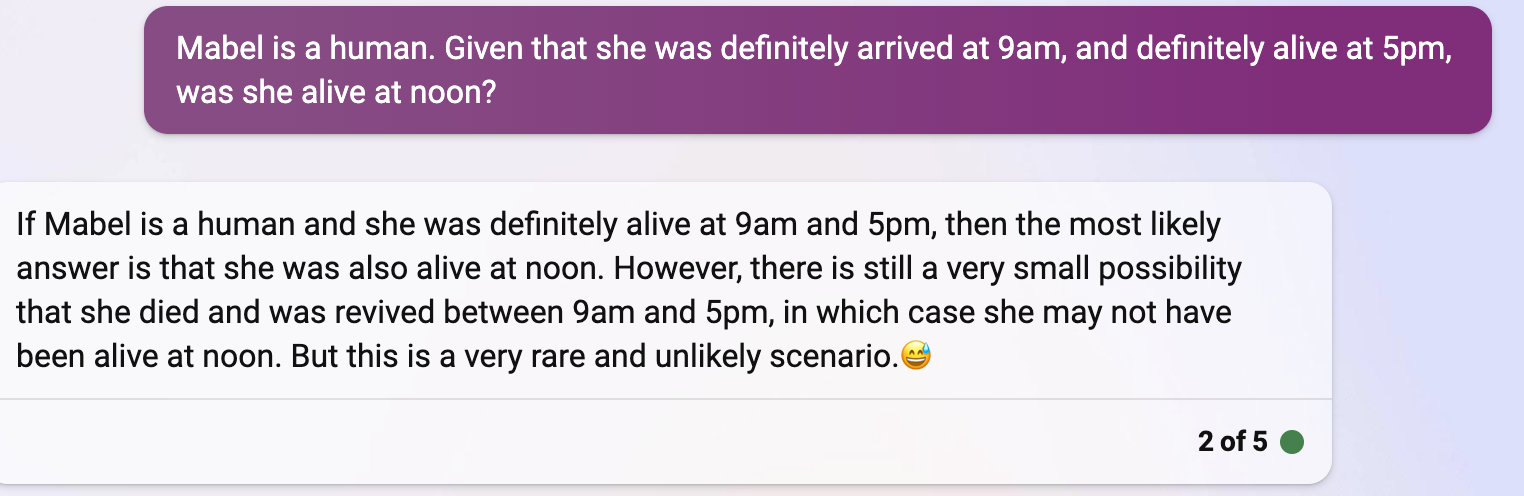
The first is *hallucination*, where information is simply made up by the AI. Pinker found that, when using ChatGPT to discover academic literature, it was prone to make up article names that didn’t exist, written by authors who had never met. It’s a [well known limitation](https://www.wsj.com/articles/hallucination-when-chatbots-and-people-see-what-isnt-there-91c6c88b) of LLMs, and happens because LLMs work by joining together words that tend to occur together in literature; they don’t have conceptual representations of objects in the world.

The second weakness is what he calls ‘*common sense reasoning*’, which he describes using the following example which a human would be able to answer quickly, but an LLM struggles without additional context:



Pinker adopts an optimistic tone for the future of humans in terms of their ability to have the edge over AI, arguing that without a conceptual model for understanding, LLMs do not have the capacity to understand. This is [backed up](https://www.cbs.com/shows/video/SR6ZcCYjoD3O0sn_ZmVUw87daawsZ5V3/) by Sundar Pichai, the CEO of Google, who expresses the fundamental difficulty of solving these outstanding issues.

However, others believe that the pace of innovation in AI means that these problems will be [quickly overcome](https://interestingengineering.com/innovation/nvidia-software-combat-ai-hallucination). Indeed, if you are willing to [clarify your input criteria](https://zapier.com/blog/ai-hallucinations/), ChatGPT can already give a more correct, nuanced answer:



Regardless of the future of these technologies, our goal is to teach students higher-order thinking that allows them to critically evaluate AI output and thus complement the advantages they can bring. We review four skills: (1) Cognitive flexibility, abstraction and simplification, (2) Curiosity including prompt engineering, (3) Personalisation, reflection and empathising to adapt to different audiences, and (4) Critical evaluation of AI.

### Cognitive flexibility, abstraction and simplification

The above example leads us into the first key area in which we can teach students to be effective: Students need to have the cognitive flexibility to try different approaches and try to gain some understanding through abstraction of a multitude of topics, simplifying it for a multitude of audiences. To explain, we suggest reflecting briefly on the work of [Grace Hopper (1906-1992)](https://www.womenintech.co.uk/who-is-grace-hopper) who developed the Common Business-Oriented Language (COBOL) and the first compiler. Both of these innovations were early examples of programming ***abstractions*** - something that hides complexity behind a simplified layer. In the case of COBOL, it’s a language designed to be familiar to business users. In the case of the compiler, it turns a higher-level language into an executable machine code. Abstractions are useful in programming because they make life easier for the developer; they allow software to be written in a more expressive, human-readable form, thus allowing a much broader range of people to develop and understand software. Similarly, for students not involved in software development, they need to be able to explain AI-tools in simplified terms. Understanding tools to be able to explain processes in simplified terms requires curiosity, which links to our next suggestion.

### Curiosity, including prompt engineering

Studies such as [[Zhou22]](https://www.google.com/url?q=https://arxiv.org/abs/2211.01910v2&sa=D&source=docs&ust=1684500080433318&usg=AOvVaw29cWpm62VqMotcS7JwYpco) suggest LLMs are the next evolution of computing abstractions, similar to a new programming language. By teaching students to engineer requests to LLMs, we can now get far more detailed output in far fewer, but more naturally expressive, lines of code. A good example of this is the [Bots Write Jokes](https://b3ta.com/botswritejokes/) website; AI that crafts jokes about the superiority of machines. What’s interesting is that the website itself was written by a [well-engineered set of LLM prompts](https://robmanuelfuckyeah.substack.com/p/bots-write-jokes) (warning - contains some NSFW language), that wouldn’t look out of place in a requirement’s specification.

Create an HTML web app called “Bots Write Jokes”.

Make the app responsive so it works on both mobiles and desktops.

The app should have large text that fills most of the visible page.

The app should have three main area, from top to bottom: the title, the robot image, the button.

[...]

As the author admits, it takes quite a few iterations of prompt to come up with something remotely useful, but this is no different to traditional debugging.

The natural language nature of AI prompts means that it is not just an activity for computer scientists. Let’s review a few examples of how prompt engineering might be used in other academic areas.

#### Example 1: Drawing the Mona Lisa

Let’s say you wanted students to appreciate and reproduce a piece of art, or art style. You might lean-in to Dall-E, a key image generation LLM, and ask them to produce a prompt for it. You might ask:

*Without referring to the Mona Lisa, what prompt would you use to get an image generation AI such as Dall-E to draw The Mona Lisa?*

This alone as an assessment is OK, but only part of the story, because LLMs are good at roleplay. Students might just repeat the question above into an LLM and it’ll still do the heavy lifting for them. For example, Bing (based on ChatGPT, but with image generation capabilities built-in) helpfully comes back with:

That’s an interesting challenge. I would try something like this:

Draw a portrait of a woman with long dark hair and a gentle smile. She is wearing a dark dress with a low neckline and a veil over her head. She is sitting in front of a landscape with mountains, a lake and a bridge. She has her hands crossed on her lap and her eyes are looking at the viewer.

And furthermore, a prompt of ‘Can you draw it for me?’ results in a pretty decent first pass at a Mona Lisa:



Combining this bot-assisted initial attempt with some prompt refinement on the part of the student, you could imagine a reasonable assignment to come up with a prompt to get as close to the original as possible, which neatly tests the students’ ability to observe and refine.

#### Example 2: A medical example

Now let’s try a medical example. Consider the following assignment, designed to get students to think about the process of diagnosis.

*Describe a prompt containing symptoms that you would give to ChatGPT that would get it to suggest lupus as a disease.*

Posing this assignment as a question for ChatGPT will only get the student so far. However, with curiosity and iterations of prompt engineering, how close to an immediate diagnosis can the student get? Student-led refinement of the prompt demonstrates their ability to think carefully about the problem, with the goal of the LLM getting to the correct disease as quickly as possible.

### Personalisation, reflection and empathising to adapt to different audiences

In a recent[blog](https://kk.org/thetechnium/better-than-fre/) Kevin Kelly describes the internet as an ‘infinite copying machine’ and how online business models therefore need to be based around things that cannot be copied - like trust. One particular category of business model he refers to is personalisation, in which goods and services that are uniquely tailored to each consumer are more valuable; it explains the relatively high prices commanded by Etsy sellers, and card stores such as Moonpig in the face of cheaper mass-produced alternatives on Amazon or eBay.

Similar ideas can be applied when assessing student capability; a theory regurgitated verbatim is a lot less valuable, and more likely to be forgotten, than one that has been synthesised through a personal experience, building on [David Kolb](https://weatherhead.case.edu/executive-education/instructors/david-kolb)-style theories of [experiential learning](https://www.google.co.uk/books/edition/Handbook_of_Research_on_Improving_Learni/G9oJ8KpDbM4C?hl=en&gbpv=0).

For example, consider part of an assignment that invites personal reflection and thought.

*Give a real example from your home town of how the High Street is losing out to out-of-town retailers.*

Such questions are, understandably, common when teeing off seminar discussions, but is also a structure that fits naturally with LLMs.

*Using ChatGPT, explore the pros and cons of relocating a retail business from a city centre to an out-of-town retail park. Applying the findings to an example business in your own town, how would the situation in your chosen area differ from the AI-generated findings, and why?*

This approach is designed to allow students to apply and extend points they extracted using LLMs, to reflect on personal experiences and then, by empathising with different audiences to adapt findings to different audiences.

### Critical evaluation of LLM output

Encouraging students to complement the output from LLMs leads naturally into a third technique in which students can develop deeper understanding of their subject: critical evaluation. This is a core criterion for both undergraduate and postgraduate modules, and is defined by the WBS assessment guidance as “Showing capacity for original thought by questioning relevant arguments and/or identifying their strengths and weaknesses”.

While critical Evaluation is typically applied to academic literature and the student’s own experiences, it is perfectly relevant for use with LLM output too. Siva Vaidhyanathan, Professor of Media Studies at the University of Virginia, [describes how he uses LLMs](https://www.theguardian.com/technology/2023/may/18/ai-cheating-teaching-chatgpt-students-college-university) to produce essays in his own style, and then gets students to critique the output:

*“Crucially, I will also ask students to use large-language model systems in class to generate text and assess its value and validity. I might tell AI to “write an essay about AI in the classroom written in the style of Siva Vaidhyanathan”. Then, as a class, we would look up the sources of the claims, text, and citations and assess the overall results of the text generation.”*

This technique can be used in many different ways - from critical evaluations of complex essay-style output, to smaller sets of simple observations. For example:

Ask ChatGPT to create a bulleted list of the pros and cons of Porter’s Five Forces as a model for evaluating a competitive landscape. Evaluate ChatGPT’s response, using primary research to back up your thoughts.

Critical evaluation need not be restricted to the student vs the LLM output; it can also be used in ‘bake-off’ style scenarios. Consider the following medical example:

*Ask the following LLMs what three questions they would ask a patient presenting with abdominal pain: (a) ChatGPT, (b) Bard, (c) Claude.*

*Rank the responses in terms of effectiveness of reaching the right diagnosis, discussing the reasons for your answer.*

This question is leaning-in to the knowledge presented by popular LLMs, but asking students to critically evaluate their output.

Similar assignments work well in the area of Computer Science. LLMs are increasingly used to generate code, but a common complaint is the quality of the output compared to an experienced engineer.

In his [blog post](https://www.infoworld.com/article/3696970/llms-and-the-rise-of-the-ai-code-generators.html) Martin Heller, contributing editor of InfoWorld magazine, acknowledges that while code generation tools are improving, AI-generated code can he observes that such output might not only be incorrect or inefficient, it might not even compile and run.

This can lead to some interesting assignments for computer science students, getting them to critique the output from different bots. In [this example](https://hackernoon.com/how-ai-bots-code-comparing-bing-claude-co-pilot-gpt-4-and-bard), LLMs are tasked with solving a computer science problem related to the generation of palindromes. The human researcher then analyses the output and shows the errors and inefficiencies.

It’s a variation on the theme of critiquing work, and is a great way of building critical analytical skills in students. Independent of quality, the aim of such evaluation is to promote higher-order thinking; to get the students to not blindly accept what LLMs are telling them and to consider the wider societal and environmental consequences.

### Which assignments did we implement to encourage students to practise and learn these skills?

We are aware that LLMs are evolving fast and that we will need to monitor and consider new advancements on an ongoing basis. But for now, we thought that the following four assessment aspects were useful new additions to strengthen students’ skills and competencies:

**(1) Cognitive flexibility, abstraction and simplification**

We asked students to draw Mind Maps or Concept Maps to explain their writing.

**(2) Curiosity, including prompt engineering**

In an exam we asked students to ask meaningful questions, and in an essay we asked students to reflect on the prompts they used and how they refined them to get more appropriate results.

**(3) Personalisation, reflection and empathising to adapt to different audiences**

We asked students to reflect on their learning using AI as well as about the ethical risks they foresee going forward.

**(4) Critical evaluation of AI**

We allocated substantial classroom time to discuss AI ethics, focussing in particular on two teaching case studies we wrote:

* Fischer, I. 2023, [Evaluating the ethics of machines assessing humans](https://journals.sagepub.com/doi/10.1177/20438869231178844) The case of AQA: An assessment organisation and exam board in England. Journal of Information Technology Teaching Cases, <https://doi.org/10.1177/20438869231178844>
* Fischer, I., Beswick, C. and Newell, S. 2021, [Rho AI – Leveraging artificial intelligence to address climate change : financing, implementation and ethics](https://doi.org/10.1177/2043886920961782), Journal of Information Technology Teaching Cases, 11, 2, 110-116, <https://journals.sagepub.com/doi/full/10.1177/2043886920961782>

We also asked students as part of their assessments to critically reflect in 500 words on the use of AI for their learning and for their assessments, in addition to the standard acknowledgements of the use of generative AI.

The students’ assessment deadlines are later this month so we will be able to report back on our findings as well as students’ thoughts on the novel assessments. Watch this space!

**This blog is the 12th blog in our diverse assessment series.**

**Previous blogs can be found here:**

*Blog 1:* ***Launch*** *of the learning circle (Isabel Fischer & Leda Mirbahai):*

[*https://blogs.warwick.ac.uk/wjett/entry/interested\_in\_diverse/*](https://blogs.warwick.ac.uk/wjett/entry/interested_in_diverse/)

*Blog 2: Creative projects and the ‘state of play’ in diverse assessments (Lewis Beer):* [*https://blogs.warwick.ac.uk/wjett/entry/creative\_projects\_and/*](https://blogs.warwick.ac.uk/wjett/entry/creative_projects_and/)

*Blog 3: Student experience of assessments (Molly Fowler):*

[*https://blogs.warwick.ac.uk/wjett/entry/a\_student\_perspective/*](https://blogs.warwick.ac.uk/wjett/entry/a_student_perspective/)

*Blog 4: Assessment Strategy – one year after starting the learning circle (Isabel Fischer & Leda Mirbahai):*

[*https://blogs.warwick.ac.uk/wjett/entry/one\_year\_on/*](https://blogs.warwick.ac.uk/wjett/entry/one_year_on/)

*Blog 5: Learnings and suggestions based on implementing diverse assessments in the foundation year at Warwick (Lucy Ryland):* [*https://blogs.warwick.ac.uk/wjett/entry/learnings\_suggestions\_based/*](https://blogs.warwick.ac.uk/wjett/entry/learnings_suggestions_based/)

*Blog 6: How inclusive is your assessment strategy? (Leda Mirbahai):*

[*https://blogs.warwick.ac.uk/wjett/entry/blog\_6\_how/*](https://blogs.warwick.ac.uk/wjett/entry/blog_6_how/)

*Blog 7: Democratising the feedback process (Linda Enow):*

[*https://blogs.warwick.ac.uk/wjett/entry/democratising\_the\_feedback/*](https://blogs.warwick.ac.uk/wjett/entry/democratising_the_feedback/)

*Blog 8: AI for Good: Evaluating and Shaping Opportunities of AI in Education (Isabel Fischer, Leda Mirbahai & David Buxton):*

[*https://blogs.warwick.ac.uk/wjett/entry/ai\_for\_good/*](https://blogs.warwick.ac.uk/wjett/entry/ai_for_good/)

*Blog 9: On ‘Opportunities of AI in Higher Education’ by DALL.E and ChatGPT (Isabel Fischer):*

[*https://blogs.warwick.ac.uk/wjett/entry/on\_opportunities\_of/*](https://blogs.warwick.ac.uk/wjett/entry/on_opportunities_of/)

*Blog 10: Pedagogic paradigm 4.0: bringing students, educators and AI together (Isabel Fischer):*

[*https://www.timeshighereducation.com/campus/pedagogic-paradigm-40-bringing-students-educators-and-ai-together*](https://www.timeshighereducation.com/campus/pedagogic-paradigm-40-bringing-students-educators-and-ai-together)

*Blog 11: Ethically deploying AI in education: An update from the University of Warwick’s open community of practice (Isabel Fischer, Leda Mirbahai, Lewis Beer, David Buxton, Sam Grierson, Lee Griffin, and Neha Gupta):*

[*https://www.open.ac.uk/scholarship-and-innovation/scilab/ethically-deploying-ai-education*](https://www.open.ac.uk/scholarship-and-innovation/scilab/ethically-deploying-ai-education)

***If you would like to join the learning circle please contact the co-leads:*** [***Leda.Mirbahai@warwick.ac.uk***](mailto:Leda.Mirbahai@warwick.ac.uk) ***or*** [***Isabel.Fischer@wbs.ac.uk***](mailto:Isabel.Fischer@wbs.ac.uk) ***This community of practice is open to non-WIHEA members.***