Lifting Weights with a Forklift:

Redesigning Assignments for Personal Growth in Anatomy and Physiology

Louis W. Kutcher
University of Cincinnati

As I write this (Spring, 2025) there's a quote circling the internet to the effect that using artificial intelligence in education is like going to the gym and watching people lift weights with a forklift; it gets the job done, but there's no personal growth.

Educators often speak about teaching critical thinking skills. This would be the "personal growth" we hope to stimulate, like the muscle growth that comes from actually lifting weights yourself. And with some justification it's fashionable to limit education to producing personal growth in our students, while implying that learning and retaining facts ("content knowledge") is of little importance. But even when theorists talk of the primacy of critical thinking, they assume that we have facts to support the critical thinking. What's not actively said is that facts must be accurate, at least to the best of our knowledge, for the critical thinking to have value.

So, then what is education? Clearly it includes gaining the mental ability, the critical thinking skills, to apply knowledge (facts) to solve real-world problems. But it also must involve the capacity to evaluate the accuracy of the facts. To call someone educated is to imply that they have both skills: the judging of facts and the application thereof.

The internet makes finding facts easier. To learn about almost any topic, type a few words into Google and a list of hundreds of websites will immediately appear. It's so common that "googling" something has become synonymous with searching for information on the internet – regardless of the search engine you use. Educators have been struggling with how to incorporate googling in teaching since Google was the shiny new toy.

Now along comes artificial intelligence (AI). Versions of AI have existed for many years; a generation of students have relied on Grammerly™ to "fix" their writing. When we speak of AI today, we usually mean *generative* AI – that is, programs that will produce novel-sounding output when given a prompt or question.

Generative AI works by predicting the most likely pattern of words to answer the query it is given. That prediction is based on a Large-Language Model (LLM) – the text or data fed into an AI program to "train" it. It's important that our students understand this: the output from an AI prompt is simply a prediction of words based on whatever information the AI has scraped from the internet. The AI engine does not comprehend either questions or answers, even if it seems to. Nor can the AI program judge the accuracy of its answers. Part of what AI does, then, is similar to any search engine; it finds facts (information) stored on the vast cloud of the internet. The other part, constructing a novel response to a prompt, is new, but it's not critical thinking.

When ChatGPT (an early AI app) was first launched, there was much concern that it would upend education by taking the place of the hard work of learning to think critically. Some of these fears have

been justified, particularly in areas that require writing long texts such as term papers. To avoid AI becoming the death knell for education – by education I mean the mental growth that occurs when students wrestle with difficult material – we must seek ways to harness the strengths of AI and warn against its dangers.

One danger with AI is when the program produces a "hallucination." Because the AI program simply predicts a response to a query, with no understanding of the underlying content, it sometimes simply makes things up out of whole cloth. And because a strength of AI is to construct coherent, well-structured sentences, it can be difficult to spot these hallucinations. It's easy to question an AI-generated image showing a dog with two heads, but you might miss that the Road Runner isn't one of the Seven Dwarfs if you're not up on Disney trivia.

Artificial intelligence programs excel at finding facts, so AI can function as a tutor… but it needs to be checked. Before AI, students would google a question and accept the top answer on the list. The advantage is that a search engine can't hallucinate; it can only point to websites that relate to the prompt. The information on the website may be incorrect, but that is subtly different from a hallucination. Ideally, students would navigate to the site, read the relevant information, and judge the validity of the source… or not.

In the classroom, I have experimented with several different assignments to help students understand the benefits and limitations of AI.

<u>In the style of…</u>

The first assignment I constructed was to have Anatomy & Physiology students put the following prompt into an AI engine:

> "Explain endochondral ossification *{basically, bone formation in our skeletal system}* in the style of…"
> where they may choose any person – real, imaginary, living or dead – to help them understand this process.

For their output, they tell who "explained" bone growth and copy their results into the worksheet. The first time I used this assignment, students answered four short prompts: what did AI get right? What did it get wrong? Was anything missing? And did they have any comments on the output? The students enjoyed the assignment, but it was unclear if they learned much content from it – they mostly agreed that everything the AI said was accurate. Later iterations have incorporated several changes:

- The topic changes to fit where we were in the semester.
- The worksheet begins with questions that probe students' understanding the basic anatomy and physiology involved.
- The questions after the AI description are more targeted. (e.g. Where did AI start the process? Did AI include certain terms? Did it explicitly describe certain relationships?)

This seems to lead to a more nuanced understanding of the AI output, but some students still miss glaring AI errors. On the instructor's side, these assignments are difficult to review since I must think about each response the AI gives, then consider if the student's analysis was sufficient or accurate. Students reported an increased understanding of the topic; however, since there were occasional errors in the AI output, one concern is that their understanding may rest on faulty facts.

Ask Me Three Times:

Another assignment that addresses using AI as a tutor is to compare the output of several iterations of a question. Since AI simply generates the most likely response to a query, the answer it gives is slightly different each time you ask.

Students should come into my class knowing the differences between ionic and covalent bonds (a topic they ideally learned in chemistry, but which we cover briefly in class). To review the topic and point to some AI challenges, their assignment is to run the following prompt through an AI engine three separate times:

> "In 100 words or less, compare and contrast ionic bonds with covalent bonds."

Adding "In 100 words or less…" was done to limit the scope of the AI response. Once the students have three versions of the same answer, they identify several common items from all three trials. They also have to identify anything that is unique to just one answer or missing from just one answer, which helps identify incorrect information. Finally, they construct their own comparison between the two types of chemical bonds. Most of the AI trials include similar information, and when they don't, the students quickly pick up on that. From the grading standpoint, this assignment is easier; I can start with the student's analysis and if I have questions *then* look back to the AI output.

These assignments lean into the ability of generative AI to scan a vast volume of information (the LLM that makes up its data set) and construct a readable answer to a prompt. The possibility of incorrect answers or hallucinations remains, which is why having students put the answers into context with limiting or clarifying questions is important. In future iterations of these assignments, I plan to add a component to emphasize the importance of vetting the answers. Students may think checking the AI output is unnecessary, but it's critical given that the AI content is not peer reviewed or fact checked.

The AI programs (sometimes called chatbots, because of their ability to interact in a conversational style) speak with confidence and authority. This leads to a universal temptation to simply accept the output, even though we know that AI can make errors and/or hallucinate. If the "weightlifting" of education is learning to evaluate information (content knowledge) and to use that knowledge in real-world scenarios, then as educators we must guide our students in both evaluating and working with facts. The internet has made finding facts easier, but AI has arguably made judging or working with them harder: it is too tempting and easy to let AI do the heavy lifting. By designing assignments that lean into the strengths of AI, and show students its pitfalls, we can target critical thinking skills as well as mastery of content.