

What is Academic Integrity in the Era of Generative Artificial intelligence?

Chris Dede, Harvard Graduate School of Education

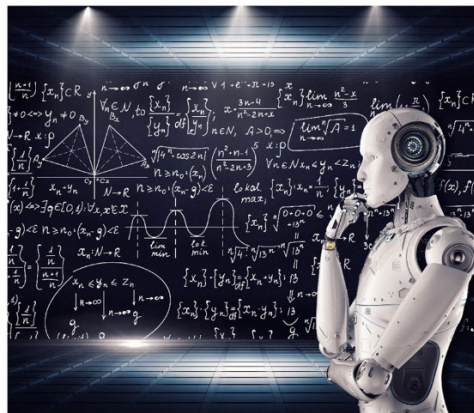
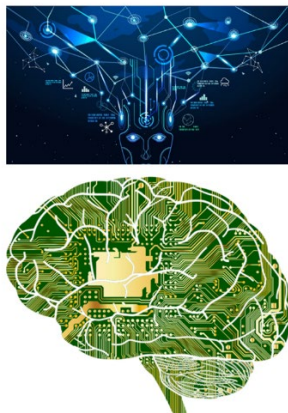


Suppose someone says, “I used generative AI to write the manuscript I turned in, but I did edit what it gave me and made sure the facts and citations were accurate. Friends tell me that generative AI produces the initial drafts of every document in their workplaces and then its output is checked by people. Why is the academic community using outdated measures of what I can perform, insisting that my artifacts be produced only by me except for limited quotations from referenced sources?”

In the new-normal of generative AI, how does one articulate the value of [academic integrity](#)? This blog presents my current response in about 2,500 words; a complete answer could fill a sizable book.

Massive amounts of misinformation are disseminated about generative AI, so the first part of my discussion clarifies what large language models (Chat-GPT and its counterparts) can do and what they cannot accomplish. The second part describes ways in which generative AI can be misused as a means of learning; unfortunately, many people are now advocating for these mistaken applications to education. The third part describes ways in which large language models (LLM), used well, can improve learning and education. I close with a plea for a robust, informed public discussion about these topics and issues raised.

What Generative AI Can and Cannot Do



I am excited about the potential of LLM to have a positive impact on education and learning. However, much of the promotion of generative AI assumes it is like “Hollywood artificial intelligence”: the evil Skynet trying to exterminate the human race in the world of the Terminator or benevolent digital deities as the next stage of evolution beyond people.

What LLM can do well is far less than Hollywood AI, but more than any type of AI has done thus far. LLM solve the natural language processing (NLP) problem. A LLM can take a textual prompt that involves paraphrases, idioms, implicit and tacit understandings, and multiple languages; process the symbols it is given; and generate a set of textual symbols in response that relate to the prompt. (Other types of representation beyond text can also be generated.) This allows fluent verbal and textual interaction between a computer program and a human being—a goal that AI has tried to achieve for more than half a century. This powerful capability of LLMs has the potential to improve learning and education in many ways.

However, because of generative AI’s ability to *seemingly* understand and rephrase language, people tend to ascribe human level cognition and affect to a LLM, believing it is more intelligent than is actually the case. This [anthropomorphism](#) is known as the [Eliza effect](#) and was discovered half a century ago with much simpler computer-based language models. Mistaken beliefs about what generative AI can do or will do in the near future are what lead to the misuses described next.

Ways in Which Learners and Anyone Who Wants to Improve *Should Not Use* Generative AI

First, in learning and improving a skill, the destination is the journey: Your written essay is not the goal, but a means to learning the skill of expressing your original thoughts in words. [Ted Chiang](#) phrases this eloquently:

...If you’re a writer, you will write a lot of unoriginal work before you write something original. And the time and effort expended on that unoriginal work isn’t wasted; on the contrary, I would suggest that it is precisely what enables you to eventually create something original. The hours spent choosing the right word and rearranging sentences to better follow one another are what teach you how meaning is conveyed by prose. Having students write essays isn’t merely a way to test their grasp of the material; it gives them experience in articulating their thoughts... Your first draft isn’t an unoriginal idea expressed clearly; it’s an original idea expressed poorly, and it is accompanied by your amorphous dissatisfaction, your awareness of the distance between what it says and what you want it to say...

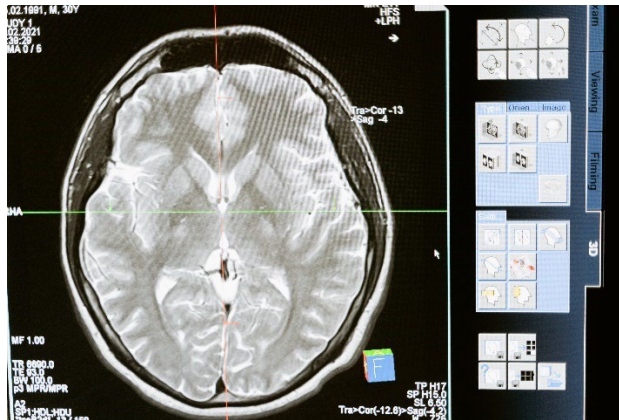
In learning, having someone/something else do your thinking is cheating yourself out of a valuable capability a good education and life experience can give you: the capacity to form an original perspective on a complex topic and express this in ways other people find meaningful. Even a simple guide like [autocomplete](#) to finish sentences can undercut original intent and thought, distorting creativity and integrity. Further, society suffers when academic institutions produce [substandard graduates](#) who can pass tests and write prompts, but don’t have vital skills.

Second, such a self-inflicted wound is made worse when generative AI is producing your essay, because LLM do not *think* in any conventional sense of the term. A LLM is an *alien* intelligence, as different from ours as anything we will find exploring the universe. When attempting to understand something novel, using analogies to familiar things can help; for me, several analogies illustrate the strengths and weaknesses of generative AI, as well as its likely limits:



A LLM is analogous to a [digital parrot](#): it can express combinations of sounds/symbols *without any understanding of these mean* or the capacity to explain how it arrived at what it is articulating. When given a prompt, a LLM must complete an utterance in response, which is formed word by word based on probabilistic predictive analytics. This is what enables a LLM to produce symbols that combine and paraphrase its inputs—this can be useful for learning but can also create serious problems.

Unlike [intelligent tutoring systems based on internal models and production rules](#), a LLM cannot provide explanations of why it chose its phrasing; it is a black-box system, not transparent. Also, its knowledge is static, so goes out of date over time unless the system is periodically refreshed and retrained.



As another analogy, a LLM is a brain (tuned digital neural networks) *without a mind*: no consciousness, metacognition, agency, senses, experiences, or [implicit knowledge of what it is like to have a biological body, a family and friends, a culture, and an ethical system with moral values](#). As such, it cannot with high competence undertake complex roles that involve these characteristics, such as a grief counselor.

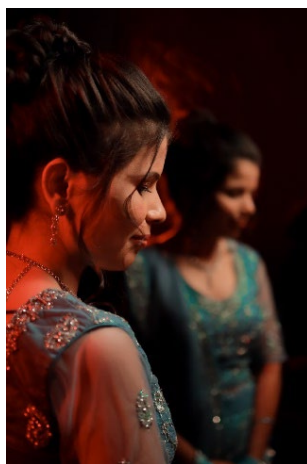
Its tuned digital neural networks are focused by a training set related to the types of prompts it is likely to receive in a particular application. But a LLM still lacks attributes that are part of human intelligence, such as reasoning, planning, thinking abstractly, comprehending complex ideas, and learning from experience. These weaknesses are very important in learning and education; teachers and mentors are effective in part both because they understand the knowledge they are communicating and because they remember what it was like to be a naïve learner.



As a result of this mindless performance, tutors based on generative AI need careful, continuous training to avoid errors and misconceptions. For example, [researchers at Stanford](#) found that, over a few months, the evolution of ChatGPT resulted in a downgrade for certain types of performance; the LLM went from correctly answering a simple math problem 98% of the time to just 2%. This type of “drift” can be caused by changes in one part of the LLM causing unexpected shifts in some other part of the tune digital neural nets. It could also be due to the introduction of false material into the parts of the worldwide web used for training; internet trolls are known to engage in this type of malicious sabotage.

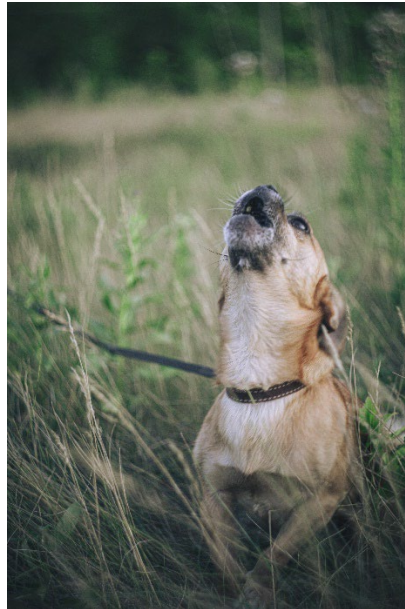
A “[walled garden](#)” approach is one way to handle this issue. The training set provides the knowledge, while the LLM is precluded from contributing other than providing a natural language interface. While its NLP is valuable, the necessity of blocking the generative AI from introducing errors and misconceptions undercuts claims made about how much the LLM itself will contribute to knowledge and compete with general human intelligence.

Overall, an important point to stress about any type of AI-based tutoring is that anything AI can teach, AI will be doing in the workplace. It is much simpler and cheaper to build an AI-based human-performance-replacement system than a tutoring system that must explain its answers. *Students who graduate knowing primarily what AI has taught them will have difficulty finding a job.*



As a final analogy, a LLM is like a distorted mirror because it is based on uncured materials assimilated from the worldwide web. Human biases and misconceptions, [uncited use of others’ intellectual property](#),

and [made-up hallucinations](#) are infused into what the Internet can offer as accurate knowledge and theory-based, empirically validated insight. For example, using LLMs as a source of academic citations is far inferior to the curated offerings a research librarian provides and, compared to using AI recommendation/search engines, includes risks such as references that do not exist and obviously useless sources selected because LLM does not comprehend them other than as a set of symbols.



But won't all these limitations disappear in the near future? I doubt that because of the fundamental constraints described above. One of the [seven deadly sins of AI predictions](#) is to assume that basic performance will lead to high competence, given enough time and resources. I offer a counter-example: Videos show that [dogs can sing](#); however, they don't sing well (by human standards, anyway) and spending billions of dollars won't improve that outcome.

Experts in AI are debating whether Intrinsic limitations of LLMs can be overcome by massive infusions of [money, energy, and digital training materials](#). In half a century of professional work, I have seen numerous [hype-cycles](#) of similar gee-whiz forecasts about AI achieving competence in sophisticated human skills, yet all those prior predictions have failed. I am skeptical about the latest optimistic assessments of *how a mindless digital parrot prone to amnesia and hallucinations will miraculously develop general intelligence comparable to humans*.

Given all this, does generative AI have legitimate promise to improve teaching, learning, and assessment? If used in ways that foster rather than undercut academic integrity, *I am excited about the potential of LLM to transform some aspects of learning and education*.

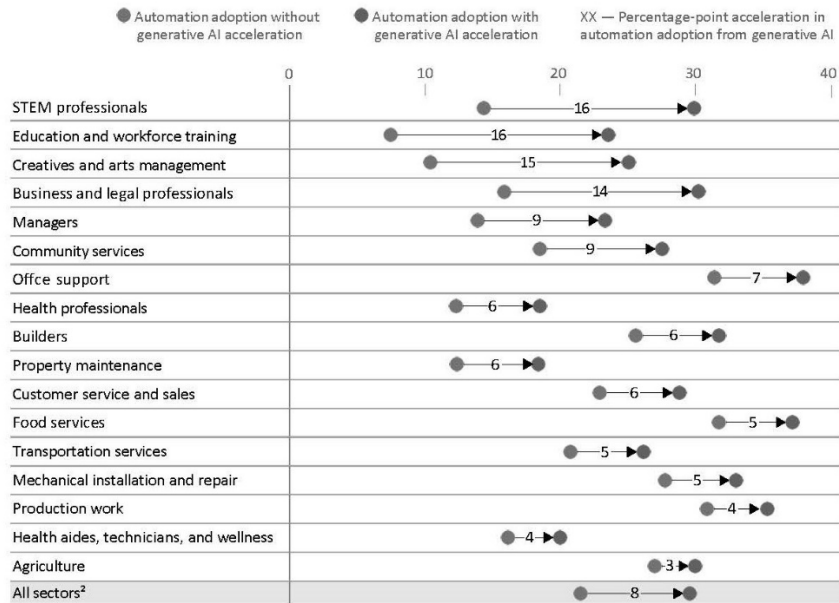
Ways Learners (and Teachers) Can Use Generative AI without Undercutting Improvement and Integrity

This is a huge topic, but I can sketch a useful framework for understanding the opportunities. In my work with the [Next Level Lab](#) at the Harvard Graduate School of Education, we study new ways of building workforce capacity for a turbulent future. While many forecasts chart an evolution of AI towards taking human jobs, more likely is a future where [AI changes the division of labor in most jobs, driving a need for workforce development to shift towards uniquely human skills](#).

Exhibit 8

With generative AI added to the picture, 30 percent of hours worked today could be automated by 2030.

Midpoint automation adoption¹ by 2030 as a share of time spent on work activities, US, %



¹Midpoint automation adoption is the average of early and late automation adoption scenarios as referenced in *The economic potential of generative AI: The next productivity frontier*, McKinsey & Company, June 2023.
²Totals are weighted by 2022 employment in each occupation.
 Source: O*NET; US Bureau of Labor Statistics; McKinsey Global Institute analysis

Specifically, AI is becoming increasingly proficient at calculation, computation, and prediction (“reckoning”) skills. As a result, we will see increased demand for human judgment skills such as decision-making under conditions of uncertainty, deliberation, ethics, and practical knowing.



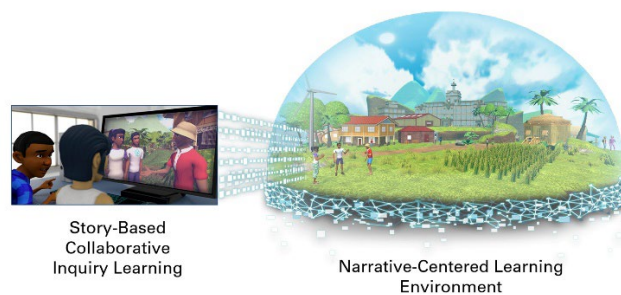
For example, In the Star Trek series, Captain Picard’s judgment, decision making, and deliberation skills are enhanced by the reckoning, computation, and calculation skills of Data, an android lacking human abilities. Together Captain Picard and Data complement one another; the synergistic combination of Picard’s judgment and Data’s reckoning provides better decision-making outcomes than the sum of their individual contributions.

In today’s workplace, people are increasingly working with AI-based partners. As an illustration, cancer specialists have access to reckoning systems that provide estimates of life expectancy and best treatment for a particular patient, based on massive amounts of data synthesized from various sources

and calculated through predictive analytics. However, healthcare workers counseling cancer patients need far more than this because real world decisions require considering quality of life versus life expectancy, tolerance for pain, personal and cultural beliefs about death, family circumstances, spiritual beliefs, and other things that no AI system can understand. The education of those workers must focus on judgment, as AI increasingly handles the reckoning; this type of learning requires academic integrity.

To learn more about the complex topic of IA, I urge that you read my co-authored Brief, [Intelligence Augmentation: Upskilling humans to complement AI](#).

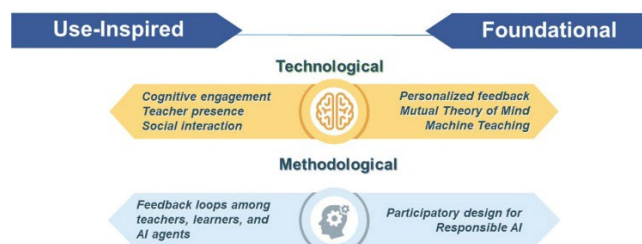
IA is a lens for understanding how AI can enhance teaching, learning, and assessment. Five National AI Institutes funded by the US National Science Foundation are currently conducting research on AI-augmented education.



One such institute, [EngageAI](#), is creating AI-driven narrative-centered learning environments to advance in STEM learning and teaching in which AI helps humans collaborate to share culturally meaningful stories.



A second Institute, [Student-AI Teaming \(iSAT\)](#), is aiming to foster deep conceptual learning via rich socio-collaborative learning experiences in which AI is a partner for the teacher.



Third, the vision of the [AI-ALOE Institute](#) is to enhance the quality of adult online education, such as tools that enhance the upskilling and reskilling of workers (including teachers); I am its Associate Director for Research.



Fourth, the goal of [Exceptional Education](#) at University of Buffalo is to develop AI solutions for children with speech and language processing challenges.



Finally, [Inclusive and Intelligent Technologies for Education \(INVITE\)](#) based at University of Illinois Urbana-Champaign (UIUC) focuses on accelerating youths' achievement in science, technology, engineering, and math (STEM) through building skills in persistence, academic resilience, and collaboration.

All of these Institutes, as well as initiatives by [other skilled researchers in AI and Education](#), are developing theory-driven, evidence-based tools and applications to empower student learning *in ways consistent with intelligence augmentation and academic integrity*. But even with these important initiatives, what can go wrong with our use of generative AI?

Three Overarching Dangers that Trouble Me



First, I am concerned that LLM will become a type of *digital duct-tape* to hold together an [obsolete industrial-era educational system](#), automating outdated instruction and assessment. Rather than focusing on using LLM to do traditional things better, we need to transform education to do better things.

For example, one can ask a LLM to suggest themes that should be in an essay or to critique a draft the student has created. While this has some value, I urge such assistance be seen as a last resort rather than initial guidance. Even a simple guide like [autocomplete](#) to finish sentences can undercut original intent and thought.



The best way to learn a skill or to improve it is initially to attempt it on your own and then get help from other people rather than from generative AI. While AI is good at reckoning, a group of peers will bring [more diverse perspectives, more creative ideas, and greater social support](#) that a chatbot can. Decades of research have validated that guided, collaborative learning-by-doing is much more effective and motivating than [conventional instruction](#) and [traditional assessment](#).



Second, some are arguing that students no longer need to learn a second language, given LLM-based translation. However, a language is [a way of thinking based on context and culture](#), and in my opinion reducing human communication to a single type of thinking conveyed by a distorted mirror is a terrible mistake. [Cultural diversity and translanguaging are strengths](#) in learning and in life.



Third, in yet another analogy, if human knowledge is sunlight, then generative AI is moonlight: The moon creates no light but reflects in a reduced way the radiance from the sun. This highlights another risk; As the Internet increasingly includes flawed products of generative AI, the material used to train LLMs is distorted and perverted in ways that decrease the quality of the “sunlight,” dumbing down civilization’s digital knowledge resources. Each successive paraphrase may increasingly [confuse the original meaning](#), like [the game of telephone](#). The analogy suggests this can be dangerous; moonlight does not support photosynthesis and does not generate enough heat to produce evaporation and rain.

Final Advice

So how do we avoid these challenges and reap the benefits of generative AI in teaching and learning? My guidance on this complex issue is the first thing I said in this blog: Don’t let anything or anyone—including me—do your thinking for you. Dialogue with many other people rather than asking a LLM or an

“expert” what the answer is. With personal integrity, [figure out your own path in learning, working, and life](#).

Photo Credits

<https://www.pexels.com/photo/man-facing-road-1248418/>

<https://www.flickr.com/photos/158301585@N08/43267970922/>

<https://openclipart.org/detail/262926/brain-computer>

<https://www.flickr.com/photos/152824664@N07/30212411048/>

<https://www.pexels.com/photo/scarlet-macaw-1618424/>

<https://www.pexels.com/photo/technology-computer-head-health-7089020/>

<https://www.pexels.com/photo/a-sinking-watercraft-in-the-sea-8017037/>

<https://www.pexels.com/photo/open-gray-wooden-door-1544422/>

<https://www.pexels.com/photo/a-reflection-of-a-woman-in-a-mirror-17099796/>

<https://www.pexels.com/photo/barking-dog-on-grass-field-9346076/>

Picard-Data Ashley Etemadi, Next Level Lab, Harvard Graduate School of Education

Graphic for Episode 12 of Silver Lining for Learning Punya Mishra, Arizona State University

<https://www.pexels.com/photo/young-female-friends-communicating-using-sign-language-in-library-7516283/>

<https://www.pexels.com/photo/photo-of-man-standing-inside-his-workshop-3811832/>

<https://www.pexels.com/photo/cirrus-clouds-577289/>