COGNITIVE RESONANCE

EDUCATION HAZARDS of GENERATIVE AI

www.CognitiveResonance.net

ABOUT THIS DOCUMENT

The *Education Hazards of Generative AI* provides a basic scientific overview of how large-language models (LLMs) work and connects this knowledge to practical implications for educators. This document is intended as a resource for teachers, principals, school district administrators, parents, students, policymakers, and anyone else thinking about using generative AI for educational purposes.

The widespread commercial deployment of LLMs, also referred to as chatbots in this document, has generated a tremendous amount of excitement, including in education. Already, teachers and administrators report using chatbots with increasing frequency. There is no shortage of hype about how LLMs will “revolutionize” education.

But although there are promising use cases for LLMs in education, there are also potential education hazards involved with using them. Chatbots are tools and, as with any tool, the failure to understand how they work may result in using them for purposes they are not well-suited for. This document highlights areas of concern where misconceptions about how LLMs function may lead to ineffective or even harmful educational practices.

The *Education Hazards of Generative AI* is intended as an introductory overview and is far from comprehensive. As a technology product, LLMs are continually being updated by the companies that deploy them, and our scientific understanding of how they function continues to evolve. That notwithstanding, and because educators are the professionals who bear the ultimate responsibility for instruction and student learning, we hope that this document is helpful in making decisions about whether or how to use generative AI in education today.

This document was co-authored by Benjamin Riley (Cognitive Resonance) and Paul Bruno (University of Illinois Urbana-Champaign). We are grateful to Amber Willis, Blake Harvard, Dan Willingham, Dylan Kane, Efrat Furst, Geoff Vaughan, Jane Rosenzweig, Jasmine Lane, Michael Pershan, Peter Greene, Sarah Oberle, Sean Trott, and Tom Mullaney for providing feedback on pre-publication drafts, along with other anonymous reviewers.

Citation:
Riley, B., & Bruno, P. (2024). *Education hazards of generative AI*.
Cognitive Resonance. https://www.cognitiveresonance.net/

What are large-language models designed to do?

Predict Text

LLMs are statistical models that take text as input and then generate text as output.[1] They are designed to address the following scenario: "Here's a fragment of text. Tell me how this fragment might go on. According to your model of the statistics of human language, what words are likely to come next?"[2]

To make their predictions, LLMs treat text as a series of *tokens*, which can be thought of as the LLM's vocabulary.[3] During training, an LLM develops the capacity to predict what text should follow a given prompt based on the frequency and relationship of the tokens included in the data it's been given. This is done through use of an *artificial neural network*, which can be thought of as a series of mathematical functions that adjust the statistical weights between tokens.[4] LLMs store these statistical relationships but not their training data – they are not search engines.[5]

LLMS ARE NOT SEARCH ENGINES

Interactions with LLMs feel conversational, and it is natural to impute human qualities to LLMs – e.g., that an LLM "knows" or "understands" something. But this can make LLMs seem more authoritative than they really are. LLMs do not determine what response would be best suited to your particular needs, and they do not necessarily produce responses that are true. It's better to think of LLMs as role-playing entities that *imitate* intelligence.[7] By design, they are intended to be helpful, and they try to do this by offering plausible responses to prompts they have been given. But their responses are often wrong.

[1]Embers of autoregression: Understanding large language models through the problem they are trained to solve

[2]Talking about large language models

[3]Tokenization in large language models, explained

[4]Large language models, explained with a minimum of math and jargon

[5]Training is not the same as chatting: ChatGPT and other LLMs don't remember everything you say

[6]Muddles about models

[7]Role-play with large language models; Imitation intelligence

EDUCATION HAZARDS POSED BY AI'S DESIGN

Lesson planning

LLMs may not correctly predict what sequence of lessons would effectively build the knowledge of students.

Generating instructional materials and assessments

LLMs may propose content that is based on common misconceptions about how students learn if those misconceptions are prevalent in the data used to train the model – for example, that students have different learning styles, or are left brained/right brained.[8]

Grading and feedback on student work

When providing feedback on essays, LLMs may not focus on the aspects of student work that are most important from a teacher's point of view. For example, LLMs may provide feedback on essays that primarily focuses on grammar or overall essay structure, even if the teacher is primarily concerned with the underlying ideas and concepts expressed by students. LLMs may also add up points or calculate percentages incorrectly if given a rubric.

Tutoring

LLMs may provide answers to students that are factually incorrect. The responses they provide may also vary depending on how they have been prompted. For example, because LLMs do not perform computations like calculators do, but instead predict likely responses to prompts, they may generate computational errors that a calculator would never make. This could result in students being misled or confused by their interactions with an LLM-based tutor.

For Administrators

LLMs may not produce content that is aligned with administrators' strategic objectives for their own schools and staff. For example, LLM-generated job descriptions or classroom observation notes may not emphasize the skills or attributes that administrators want in their staff. Such LLM-generated content also may not align with applicable laws, regulations, or collective bargaining agreements, which could result in legal liability.

[8]The science of learning

Do large-language models learn the way that humans do?

No

The most powerful LLMs are trained on a huge amount of data. But LLMs are not continuously learning from their interactions with humans, and they do not distinguish useful data from misleading data during their training.

To produce their output, LLMs are trained on data that has been produced by humans, mostly from digital sources found on the Internet, including Wikipedia, academic articles, news stories, books, and computer code.[9] Although the data used to train LLMs is vast, it does not encompass knowledge that is not encoded digitally. Further, the models available in the United States are primarily trained in English, and on data that academics label "WEIRD": Western, Educated, Industrialized, Rich, and Democratic. LLMs are thus exposed to a biased sample of cultural practices and values.[10]

The commercial companies that make the largest and most well-known LLMs no longer disclose the precise data that they use to train their models. There are pending lawsuits against some of these companies for violating copyright and other laws protecting intellectual property.[11]

After being trained on large data sets, LLMs are *fine-tuned* by the commercial companies that produce them. This process can include testing the model with human users and asking them to provide feedback on the model's output, and then using this to adjust the statistical weights the model uses to make text predictions[12] We do not know precisely how individual companies fine tune their LLMs. Numerous scholars have called attention to existing and potential biases that presently exist in LLMs that result from how they are trained.[13]

LLMS ARE EXPOSED TO A BIASED SAMPLE OF CULTURAL PRACTICES AND VALUES

[9]A philosophical introduction to language models – part I: Continuity with classic debates

[10]GPT-4 is WEIRD – What should we do about it?; Which humans?

[11]Copyright safety for generative AI

[12]A philosophical introduction to language models

[13]On hate scaling laws for data-swamps; On the dangers of stochastic parrots: Can language models be too big?

EDUCATION HAZARDS POSED BY AI'S TRAINING

Lesson planning

Given that many instructional materials found online are low-quality, LLMs may not have been trained on high-quality lesson plans or on lesson plans aligned with specific content standards.[14]

Generating instructional materials and assessments

The materials LLMs generate may not align to the needs of culturally or linguistically diverse students if the LLM training data does not include text from these students' communities. Data on the Internet is not representative of knowledge globally. LLM-based materials may also include inaccurate or false information that's prevalent in the data used to train the LLM. This could include misinformation about content (e.g., the biological processes of evolution) or misconceptions how students learn (e.g., that lessons need to be differentiated for "visual learners").[15]

Grading essays and providing feedback

LLMs may not recognize student creativity if the student's work does not align to the data they have been trained on

Tutoring

In general, LLMs do not learn from their interactions with students. The LLM's capabilities are almost entirely derived from its training data. This means that LLMs may not adapt to the specific needs of the students they are tutoring.

For administrators and policymakers

Administrators should emphasize to teachers and others using LLM-created materials that educators are responsible for the validity and usefulness of the materials they choose to use, including in personnel evaluation contexts. Likewise, if school administrators encourage or mandate the use of LLM-based tools by teachers and other educators, administrators should be held responsible for the validity and usefulness of those resources.

[14]The supplemental curriculum bazaar: Is what's online any good?

[15]Ask the cognitive scientist: Does tailoring instruction to "learning styles" help students learn?

Can large-language models reason?

Not like humans

With careful prompting, LLMs can produce text that appears to involve higher-order thinking skills, but it's better to see LLMs as skilled at *pattern matching*. Be wary of any definitive statement suggesting that LLMs are currently capable of reasoning the way that humans do – the evidence does not support such strong claims right now.

An important discovery with LLMs is that the quality of their output improves when they are prompted in specific ways. One method is called *chain-of-thought* prompting where LLMs are given an example of a correctly solved problem first and then asked to solve a similar problem.[16] This is similar to using worked examples to help students understand a concept before asking them to solve a related problem.[17] Another method is simply to ask LLMs to "think step by step," which can also improve performance.[18]

Because LLMs can produce text related to complex ideas after careful prompting, some claim that they are already as intelligent as humans.[19] But across a wide range of novel tasks that fall outside the sort of data that LLMs are trained upon, LLMs perform significantly worse than humans.[20] At present, the weight of the evidence suggests LLMs produce their output predominantly by pattern matching inputted text to data they have been trained upon, rather than reasoning in human fashion.[21]

LLMS PRODUCE THEIR OUTPUT BY PATTERN MATCHING, NOT REASONING

[16]Chain-of-Thought prompting elicits reasoning in large language models

[17]How to make worked examples work

[18]Large language models are zero-shot reasoners

[19]Why the godfather of A.I. fears what he's built

[20]Language models don't always say what they think: Unfaithful explanations in chain-of-thought prompting; Comparing humans, GPT-4, and GPT-4V on abstraction and reasoning tasks; Embers of autoregression; The reversal curse: LLMs trained on "A is B" fail to learn "B is A"

[21]Can large language models reason?; Language models don't always say what they think

EDUCATION HAZARDS POSED BY AI'S LACK OF REASONING

Generating lesson plans, instructional materials, and assessments

Educators should provide LLMs with examples of high-quality content when prompting them to produce new material. The more complex the topic, the more risk there is that LLMs will produce plausible but factually incorrect materials.

Grading essays and providing feedback

Educators who choose to use LLMs to give feedback to students should provide LLMs with a range of examples of student work aligned to scoring rubrics, as well as a range of human feedback. Results should be carefully monitored by educators.

Tutoring

Students should understand that providing more context on their thinking when prompting an LLM may help improve the quality of its response. Teachers should review LLM output related to complex concepts. Rather than assuming LLMs can reason about a prompt, the user should model the type of reasoning they are looking for. LLMs may produce explanations on complex ideas that sound plausible to students but are not logical or coherent.

For administrators and policymakers

Professional development for educators on how to use LLMs should not focus exclusively or even primarily on "prompt engineering." Instead, to the extent time and resources are dedicated to LLM usage at all, they may be better spent by first building general knowledge of how LLMs function, and then rigorously evaluating the impact they have on student learning or other important outcomes.

Does AI make the content we teach in schools obsolete?

No

A bedrock principle of cognitive science is that humans understand new ideas based on ideas we already know, i.e., knowledge that is stored in our long-term memory.[22] Knowledge cannot be outsourced to AI, and students who have not built a broad base of knowledge will not be able to make best of use this new technology. Educators should continue to focus on building student knowledge across all subjects.

Not long ago, it was common to hear some educators ask, "why teach it if students can Google it?" With the commercial deployment of chatbots, some educators ask, "why teach it if students can have generative AI do it?"

Cognitive science provides an unequivocal answer to these questions: Students need to develop a broad base of knowledge – in their heads – to learn new ideas and navigate the world they experience.[23] The fact that chatbots can generate essays, summarize ideas, and create other things is an impressive technological achievement, but it does not affect how our minds work.

What's more, effective use of LLMs requires the user to possess existing background knowledge and expertise.[24] Applying this knowledge when interacting with an LLM can lead to productive, co-created output, but there are no shortcuts – if students lack this knowledge, their ability to make use of this technology will be severely limited.[25]

EFFECTIVE USE OF LLMS REQUIRES HUMAN KNOWLEDGE AND EXPERTISE

[22]The science of learning

[23]How knowledge helps: It speeds and strengthens reading comprehension, learning—and thinking

[24]Transcript: Ezra Klein interviews Ethan Mollick

[25]Learning that doesn't stick

EDUCATION HAZARDS POSED BY AI SUPPLANTING KNOWLEDGE AND SKILLS

Reading

LLMs can quickly provide summaries of complex texts, and both educators and students will be tempted to use them for this purpose. But reading complex ideas and thinking about their implications is vital to building knowledge that can be applied in other contexts. Educators should avoid using LLM-created materials that create less-challenging texts for certain students – differentiating texts this way is inequitable and may cause long-term educational harm.

Writing

Educators should be particularly careful regarding the use of AI for writing tasks. In many cases, the purpose of a writing assignment is to make students think effortfully and develop their ideas through the writing process. When students rely on chatbots to assist their writing, they can miss opportunities to learn how to think critically, to assess ideas, and to consider alternate viewpoints.[26] But educators should not use AI detection tools to prevent students from using chatbots to write their essays – such tools are unreliable, and often are discriminatory against non-native English speakers.[27]

Science, history, and other content subjects

LLMs are known to "hallucinate," that is, they can generate text that sounds plausible but does not accord with the truth.[28] For subjects where there is a great deal of factually accurate text available on the Internet, chatbots may produce output that we would recognize as true. But if information on a subject is sparse, or if there is misinformation on the topic online, LLMs may produce output that isn't true. Educators cannot rely on LLMs to be factually accurate, and thus will need to fact-check any materials they create.

Mathematics

Unlike computers or calculators, LLMs generally do not solve math problems by applying formal rules of mathematics. Instead, they treat math-related prompts as text, and then predict what text to produce as output. This can lead to outputs that "sound right" but sometimes include mathematical or logical errors that students may miss. Educators will need to monitor LLM output carefully if used for math instruction.

Computer science

The implications of LLMs for computer science education are highly uncertain. The commercial deployment of LLMs does not necessarily imply that students need to focus on coding to effectively participate in society and the economy. For instance, AI tools may be able to do some of the work currently performed by human software engineers, reducing demand for coding skills in the labor market. At the same time, students may benefit from learning about the computer science underlying LLMs – such as the information contained in this document – so that they can use LLM-related tools more effectively.

For school administrators and policymakers

Education leaders should avoid oversimplifications such "if AI can do this, we don't need to teach it anymore" or "AI will be a part of jobs of the future so we should let students use AI today." Likewise, leaders should refrain from funding professional development to teachers that suggests AI, or any other technology, supplants the need for educators to build student knowledge across the traditional range of subjects in school. Think carefully about the value of a skill before eliminating it from standards or curriculum.

[26] When the friction is the point

[27] GPT detectors are biased against non-native English writers

[28] Hallucinations, errors, and dreams; What teachers should know about the ELIZA effect

Will large-language models become smarter than humans?

No one knows

The claim that LLMs are on the verge of matching human cognitive ability in the near-term, or that they will soon be "smarter" or more effective than human educators, is not well-supported by existing scientific evidence. Educators should engage with LLMs as they currently function and be skeptical of speculative claims about the future capabilities of this technology.

Some of the commercial companies that produce LLMs say they are working to create "artificial general intelligence," AI systems that are as intelligent (or more intelligent) than humans.[29] Whether this is possible is the subject of heated debate in the AI research community. On the one hand, the rate of performance improvement of LLMs might be slowing down, and most models seem to be converging on the same level of capabilities.[30] There is growing evidence that large-language models are limited to producing output related to their training data and cannot reliably "generalize" to novel situations they haven't been trained on, and some scholars suggest they never will be able to do so.[31]

On the other hand, some argue that adding new data and more computing power to LLMs will continue to drive exponential improvement in their capabilities. And others believe that as future generative AI models incorporate an increasingly wide range of input beyond just human language, human-like intelligence will result.[32]

But no one can predict what the future holds. Be skeptical of those who claim otherwise.

EDUCATION HAZARDS POSED BY SPECULATION ABOUT AI'S FUTURE

For Teachers and Students

Educators and students should not assume LLMs will continually improve – the better approach is to engage with this technology as it exists now to explore its strengths and limitations. Engaging with AI does not necessarily mean integrating it into instructional practice.

For school administrators and policymakers

Leaders should not invest time and resources to incorporate AI in schools based on assumptions about what the future will bring. Nor should they drastically alter curricula to prepare students for an "AI world." We simply do not know what such a world will look like or what it will require of future citizens.

[29] Planning for AGI and beyond

[30] AI scaling myths; Two years later, deep learning is still faced with the same fundamental challenges

[31] Reasoning or reciting? Exploring the capabilities and limitations of language models through counterfactual tasks; The generative AI paradox: 'What it can create, it may not understand'; LLMs have failed every single benchmark and experiment focused on generalization, since their inception; AI and the Everything in the Whole Wide World Benchmark

[32] Will scaling work?