

**TO DO:**

- Read the following article published in *Science News* and understand the contents.
  - Go through the discussion questions on pp. 5-6, try to have some ideas, and prepare for the discussion that will take place during the Trial Lesson.
- 

## **How ChatGPT and similar AI will disrupt education Teachers are concerned about cheating and inaccurate information**

Source: *ScienceNews*  
April 12, 2023, 7:00 am  
By Kathryn Hulick



“We need to talk,” Brett Vogelsinger said. A student had just asked for feedback on an essay. One paragraph stood out. Vogelsinger, a ninth grade English teacher in Doylestown, Pa., realized that the student hadn’t written the piece himself. He had used ChatGPT.

The artificial intelligence tool, made available for free late last year by the company OpenAI, can reply to simple prompts and generate essays and stories. It can also write code.

Within a week, it had more than a million users. As of early 2023, Microsoft planned to invest \$10 billion into OpenAI, and OpenAI’s value had been put at \$29 billion, more than double what it was in 2021.

It’s no wonder other tech companies have been racing to put out competing tools. Anthropic, an AI company founded by former OpenAI employees, is testing a new chatbot called Claude. Google launched Bard in early February, and the Chinese search company Baidu released Ernie Bot in March.

A lot of people have been using ChatGPT out of curiosity or for entertainment. I asked it to invent a silly excuse for not doing homework in the style of a medieval proclamation. In less than a second, it offered me: “Hark! Thy servant was beset by a horde of mischievous leprechauns, who didst steal mine quill and parchment, rendering me unable to complete mine homework.”

But students can also use it to cheat. ChatGPT marks the beginning of a new wave of AI, a wave that’s poised to disrupt education.

When Stanford University’s student-run newspaper polled students at the university, 17 percent said they had used ChatGPT on assignments or exams at the end of 2022. Some admitted to submitting the chatbot’s writing as their own. For now, these students and others are probably getting away with it. That’s because ChatGPT often does an excellent job.

“It can outperform a lot of middle school kids,” Vogelsinger says. He might not have known his student had used it, except for one thing: “He copied and pasted the prompt.”

The essay was still a work in progress, so Vogelsinger didn’t see it as cheating. Instead, he saw an opportunity. Now, the student and AI are working together. ChatGPT is helping the student with his writing and research skills.

“[We’re] color-coding,” Vogelsinger says. The parts the student writes are in green. The parts from ChatGPT are in blue. Vogelsinger is helping the student pick and choose a few sentences from the AI to expand on — and allowing other students to collaborate with the tool as well. Most aren’t turning to it regularly, but a few kids really like it. Vogelsinger thinks the tool has helped them focus

their ideas and get started.

This story had a happy ending. But at many schools and universities, educators are struggling with how to handle ChatGPT and other AI tools.

In early January, New York City public schools banned ChatGPT on their devices and networks. Educators were worried that students who turned to it wouldn't learn critical-thinking and problem-solving skills. They also were concerned that the tool's answers might not be accurate or safe. Many other school systems in the United States and around the world have imposed similar bans.

Keith Schwarz, who teaches computer science at Stanford, said he had "switched back to pencil-and-paper exams," so students couldn't use ChatGPT, according to the Stanford Daily.

Yet ChatGPT and its kin could also be a great service to learners everywhere. Like calculators for math or Google for facts, AI can make writing that often takes time and effort much faster. With these tools, anyone can generate well-formed sentences and paragraphs. How could this change the way we teach and learn?

### ***The good, bad and weird of ChatGPT***

ChatGPT has wowed its users. "It's so much more realistic than I thought a robot could be," says Avani Rao, a sophomore in high school in California. She hasn't used the bot to do homework. But for fun, she's prompted it to say creative or silly things. She asked it to explain addition, for instance, in the voice of an evil villain.

Given how well it performs, there are plenty of ways that ChatGPT could level the playing field for students and others working in a second language or struggling with composing sentences. Since ChatGPT generates new, original material, its text is not technically plagiarism.

Students could use ChatGPT like a coach to help improve their writing and grammar, or even to explain subjects they find challenging. "It really will tutor you," says Vogelsinger, who had one student come to him excited that ChatGPT had clearly outlined a concept from science class.

Educators could use ChatGPT to help generate lesson plans, activities or assessments — perhaps even personalized to address the needs or goals of specific students.

Xiaoming Zhai, an expert in science education at the University of Georgia in Athens, tested ChatGPT to see if it could write an academic paper. He was impressed with how easy it was to summarize knowledge and generate good writing using the tool. "It's really amazing," he says.

All of this sounds wonderful, but really big problems exist. Most worrying, ChatGPT and other similar tools can often get things very wrong. They don't pull facts from databases. Rather, they are trained to generate new text that sounds natural. They remix language without understanding it, which can lead to glaring mistakes.

The news website CNET came under fire earlier this year for using AI to churn out dozens of articles, many of them packed with errors. In an early advertisement for the Bard chatbot, it made a factual error about the James Webb Space Telescope, incorrectly claiming that it had taken the very first picture of an exoplanet. And ChatGPT said in a conversation posted on Twitter that the fastest marine mammal was the peregrine falcon. A falcon, of course, is a bird and doesn't live in the ocean.

ChatGPT is "confidently wrong," says Casey Fiesler, an expert in the ethics of technology at the University of Colorado Boulder. "There are mistakes and bad information." She has made multiple TikTok videos about the pitfalls of ChatGPT.

Most of ChatGPT's training data come from before September 2021, and it does not provide sources for its information. If asked for sources, it makes them up, Fiesler revealed in one video. Zhai, who sees the tool as an assistant, discovered the exact same thing. When he asked ChatGPT for citations,

it gave him sources that looked correct. But they didn't actually exist.

### ***How ChatGPT works***

ChatGPT's mistakes make sense if you know how it works. "It doesn't reason. It doesn't have ideas. It doesn't have thoughts," explains Emily M. Bender, a computational linguist at the University of Washington in Seattle.

ChatGPT was developed using at least two types of machine learning. The primary type is a large language model based on an artificial neural network. Loosely inspired by how neurons in the brain interact, this computing architecture finds statistical patterns in vast amounts of data.

A language model learns to predict what words will come next in a sentence or phrase by churning through vast amounts of text. It places words and phrases into a multidimensional map that represents their relationships to one another. Words that tend to come together, like peanut butter and jelly, end up closer together in this map.

The size of an artificial neural network is measured in parameters. These internal values get tweaked as the model learns. In 2020, OpenAI released GPT-3. At the time, it was the biggest language model ever, containing 175 billion parameters. It had trained on text from the internet as well as digitized books and academic journals. Training text also included transcripts of dialog, essays, exams and more, says Sasha Luccioni, a Montreal-based researcher at Hugging Face, a company that builds AI tools.

OpenAI improved upon GPT-3 to create GPT-3.5. In early 2022, the company released a fine-tuned version of GPT-3.5 called InstructGPT. This time, OpenAI added a new type of machine learning. Called reinforcement learning with human feedback, it puts people into the training process.

These workers check the AI's output. Responses that people like get rewarded. Human feedback can also help reduce hurtful, biased or inappropriate responses. This fine-tuned language model powers freely available ChatGPT. As of March, paying users receive answers powered by GPT-4, a bigger language model.

During ChatGPT's development, OpenAI added extra safety rules to the model. It will refuse to answer certain sensitive prompts or provide harmful information. But this step raises another issue: Whose values are programmed into the bot, including what it is — or is not — allowed to talk about?

OpenAI is not offering exact details about how it developed and trained ChatGPT. The company has not released its code or training data. This disappoints Luccioni because it means the tool can't benefit from the perspectives of the larger AI community. "I'd like to know how it works so I can understand how to make it better," she says.

When asked to comment on this story, OpenAI provided a statement from an unnamed spokesperson. "We made ChatGPT available as a research preview to learn from real-world use, which we believe is a critical part of developing and deploying capable, safe AI systems," the statement said. "We are constantly incorporating feedback and lessons learned." Indeed, some experimenters have gotten the bot to say biased or inappropriate things despite the safety rules. OpenAI has been patching the tool as these problems come up.

ChatGPT is not a finished product. OpenAI needs data from the real world. The people who are using it are the guinea pigs. Notes Bender: "You are working for OpenAI for free."

### ***ChatGPT's academic performance***

How good is ChatGPT in an academic setting? Catherine Gao, a doctor and medical researcher at Northwestern University's Feinberg School of Medicine in Chicago, is part of one team of researchers that is putting the tool to the test.

Gao and her colleagues gathered 50 real abstracts from research papers in medical journals and

then, after providing the titles of the papers and the journal names, asked ChatGPT to generate 50 fake abstracts. The team asked people familiar with reading and writing these types of research papers to identify which were which.

“I was surprised by how realistic and convincing the generated abstracts were,” Gao says. The reviewers mistook roughly one-third of the AI-generated abstracts as human-generated.

In another study, Will Yeadon and colleagues tested whether AI tools could pass a college exam. Yeadon, a physics instructor at Durham University in England, picked an exam from a course that he teaches. The test asks students to write five short essays about physics and its history. Students have an average score of 71 percent, which he says is equivalent to an A in the United States.

Yeadon used the tool davinci-003, a close cousin of ChatGPT. It generated 10 sets of exam answers. Then Yeadon and four other teachers graded the answers using their typical standards. The AI also scored an average of 71 percent. Unlike the human students, though, it had no very low or very high marks. It consistently wrote well, but not excellently. For students who regularly get bad grades in writing, Yeadon says, it “will write a better essay than you.”

These graders knew they were looking at AI work. In a follow-up study, Yeadon plans to use work from the AI and students and not tell the graders whose is whose.

### ***Tools to check for cheating***

When it’s unclear whether ChatGPT wrote something or not, other AI tools may help. These tools typically train on AI-generated text and sometimes human-generated text as well. They can tell you how likely it is that text was composed by an AI. Many of the existing tools were trained on older language models, but developers are working quickly to put out new, improved tools.

A company called Originality.ai sells access to a tool that trained on GPT-3. Founder Jon Gillham says that in a test of 10,000 samples of texts composed by models based on GPT-3, the tool tagged 94 percent of them correctly as AI-generated. When ChatGPT came out, his team tested a smaller set of 20 samples. Each only 500 words in length, these had been created by ChatGPT and other models based on GPT-3 and GPT-3.5. Here, Gillham says, the tool “tagged all of them as AI-generated. And it was 99 percent confident, on average.”

In late January 2023, OpenAI released its own free tool for spotting AI writing, cautioning that the tool was “not fully reliable.” The company is working to add watermarks to its AI text, which would tag the output as machine-generated, but doesn’t give details on how. Gillham describes one possible approach: Whenever it generates text, the AI ranks many different possible words for each position. If its developers told it to always choose the word ranked in third place rather than first place at specific points in its output, those words could act as a fingerprint, he says.

As AI writing tools improve, the tools to sniff them out will need to improve as well. Eventually, some sort of watermark might be the only way to sort out true authorship.

### ***ChatGPT and the future of writing***

There’s no doubt we will soon have to adjust to a world in which computers can write for us. But educators have made these sorts of adjustments before. As high school student Rao points out, Google was once seen as a threat to education because it made it possible to look up facts instantly. Teachers adapted by coming up with teaching and testing materials that don’t depend as heavily on memorization.

Now that AI can generate essays and stories, teachers may once again have to rethink how they teach and test. Rao says: “We might have to shift our point of view about what’s cheating and what isn’t.”

Some teachers will prevent students from using AI by limiting access to technology. Right now, Vogelsinger says, teachers regularly ask students to write out answers or essays at home. “I think those assignments will have to change,” he says. But he hopes that doesn’t mean kids do less writing.

Teaching students to write without AI's help will remain essential, agrees Zhai. That's because "we really care about a student's thinking," he stresses. And writing is a great way to demonstrate thinking. Though ChatGPT can help a student organize their thoughts, it can't think for them, he says.

Kids still learn to do basic math even though they have calculators (which are often on the phones they never leave home without), Zhai acknowledges. Once students have learned basic math, they can lean on a calculator for help with more complex problems.

In the same way, once students have learned to compose their thoughts, they could turn to a tool like ChatGPT for assistance with crafting an essay or story. Vogelsinger doesn't expect writing classes to become editing classes, where students brush up AI content. He instead imagines students doing prewriting or brainstorming, then using AI to generate parts of a draft, and working back and forth to revise and refine from there.

Though he's overwhelmed about the prospect of having to adapt his teaching to another new technology, he says he is "having fun" figuring out how to navigate the new tech with his students.

Rao doesn't see AI ever replacing stories and other texts generated by humans. Why? "The reason those things exist is not only because we want to read it but because we want to write it," she says. People will always want to make their voices heard.

(2,609 words)

**Editor's Note:**

This story was updated May 10, 2023, to correct that Jon Gillham's AI-detection tool was 99 percent confident in identifying AI-generated text (not that the tool correctly identified 97 percent of the AI-generated text). A version of this article appears in the April 8, 2023 issue of Science News.

**Citations:**

C.A. Gao et al. Comparing scientific abstracts generated by ChatGPT to original abstracts using an artificial intelligence output detector, plagiarism detector, and blinded human reviewers. Preprint published online December 27, 2022. doi: 10.1101/2022.12.23.521610.

W. Yeadon et al. The Death of the Short-Form Physics Essay in the Coming AI Revolution. Preprint published online December 2022. doi: 10.48550/arXiv.2212.11661.

**Discussion Questions:**

Read the following questions and prepare for the discussions during the Trial Lesson.

- *Basic Questions:* The questions that can be answered by reading the article.
- *Advanced Questions:* The questions that will help you prepare your opinions for the discussions. Also, go online and investigate to develop your ideas.

*Note: You do not have to answer all questions. Choose some of them and think about the topic of the reading material.*

**Basic Questions:**

1. According to the author, what are some of the reasons that people have been trying ChatGPT?
2. According to educator Brett Vogelsinger, how does ChatGPT perform on writing assignments compared to middle school students?
3. Casey Feisler says that ChatGPT is "confidently wrong"? What is meant by that expression?
4. What was the New York City Public School's response to ChatGPT?
5. ChatGPT can make mistakes. What examples does the article provide of the kind of mistakes that ChatGPT can make?
6. The news site CNET faced severe criticism because it used AI to produce a great number of news articles full of mistakes and wrong information. What were the causes of the errors, and how would it be possible to fix problems like this?



7. What unit is used to measure the size of the neural networks in AI language models such as ChatGPT?
8. In a test comparing human-generated article abstracts with AI-generated article academic abstracts, how did ChatGPT fare?
9. When Will Yeadon compared the ChatGPT-generated test exam results with those of human subjects, he found that the average scores were virtually the same. What interesting difference did he note?
10. Brett Vogelsinger noted that his students were using ChatGPT in his classes. How did he react? Why did he make the decision that he did, and do you think that he made the appropriate decision?

**Advanced Questions:**

1. If ChatGPT can make mistakes, how can we check the accuracy of its suggestions?
2. Is it possible to trick ChatGPT into asking questions it cannot answer? Do you have any examples of questions that are too complicated for ChatGPT?
3. Some universities are returning to 'pen and paper' exams to prevent students from cheating: is this a good idea? If not, what are some ways to ensure that students do not utilize text generating AI programs like ChatGPT during exams or when writing papers?
4. Artificial intelligence is widely used in everyday life beyond the classroom. Can you provide examples of AI that we encounter in our everyday lives?
5. What are some ways to ensure that AI tools enhance student performance and understanding?
6. Do you think we need rules for the usage of AI tools to guarantee using them in the right way? If so, who would make and enforce such rules for the usage of AI tools?
7. According to Emily M. Bender, a computational linguist, in what ways does ChatGPT differ from human intelligence?
8. In your opinion, how likely is it that a Japanese company will develop text generating AI applications like ChatGPT?
9. In your view, do the advantages of using AI for educational purposes outweigh the disadvantages?
10. The article says that because ChatGPT generates "new, original material" it "is not technically plagiarism." Do you agree with that assessment? What is your definition of plagiarism?
11. Why does ChatGPT allow people to use its product for free? What does Chat GPT gain from allowing people to use its product?
12. The authors use the analogy: ChatGPT is to writing what the calculator is to math. Explain why that is an appropriate analogy (or not).
13. Avani Rao, a high school student from California, believes that ChatGPT will never fully replace actual composition because people "want to make their voices heard." What does Rao mean by that statement and do you agree with her assessment?
14. What are the most important elements of a liberal arts education? How could technology like AI be used to enhance and 'support' a traditional liberal arts education?

## Appendix:

When prompted, ChatGPT can craft answers that sound surprisingly like those from a student. We asked middle school and high school students from across the country, all participants in our Science News Learning education program, to answer some basic science questions in two sentences or less. The examples throughout the story compare how students responded with how ChatGPT responded when asked to answer the question at the same grade level.

Here are some sample answers of students and those of generated by ChatGPT:

### **What effect do greenhouse gases have on the Earth?**

*Agnes B.* Grade 11, Harbor City International School, Minn.

Greenhouse gases effectively trap heat from dissipating out of the atmosphere, increasing the amount of heat that remains near Earth in the troposphere.

*ChatGPT*

Greenhouse gases trap heat in the Earth's atmosphere, causing the planet to warm up and leading to climate change and its associated impacts like sea level rise, more frequent extreme weather events and shifts in ecosystems.

### **Define the term biodiversity.**

*Dante A.* Grade 10, Clark Magnet High School, Calif.

Biodiversity refers to the variety of living species and ecosystems that exist in a particular region or on the planet as a whole. It encompasses the different genes, species and ecosystems that make up the natural world and the relationships between them.

*ChatGPT*

Biodiversity refers to the variety of living organisms that inhabit the Earth and the ecosystems they form. It includes the diversity of species, genes and ecosystems, and is important for maintaining the balance of nature and sustaining life on our planet.

### **What are black holes and where are they found?**

*Althea C.* Grade 11, Waimea High School, Hawaii

A black hole is a place in space where gravity is so strong that nothing, not even light, may come out.

*ChatGPT*

Black holes are extremely dense regions in space where the gravity is so strong that not even light can escape, and they are found throughout the universe.

### **What is heat?**

*Precious A.* Grade 6, Canyon Day Junior High School, Ariz.

Heat is the transfer of kinetic energy from one medium or object to another, or from an energy source to a medium or object through radiation, conduction and convection.

*ChatGPT*

Heat is a type of energy that makes things warmer. It can be produced by burning something or through electricity.

### **What is DNA and how is it organized?**

*Luke M.* Grade 8, Eastern York Middle School, Pa.

DNA, or deoxyribonucleic acid, is kept inside the cells of living things, where it holds instructions for the genetics of the organism it is inhabiting.

*ChatGPT*

DNA is like a set of instructions that tells our cells what to do. It's organized into structures called chromosomes, which contain all of the DNA in a cell.