Beyond the Hype: Teaching Students to Recognize the Hidden Dangers of AI

Elizabeth A. Radday

Abstract

As generative AI becomes ubiquitous in classrooms and daily life, students need to learn and practice critical thinking. This article explores four of the most prevalent dangers of AI: bias, hallucinations/misinformation, deep fakes, and emotional attachment to chatbots. Educators need to not only be aware of these pitfalls of AI but also need to prepare students to look for and face these dangers with their human intelligence. Ethical, informed AI use begins with awareness, and that awareness must start in K-12 schools.

Keywords: AI Literacy • Digital Media Literacy • AI in K-12 Education.

Artificial Intelligence (AI) has made an indelible mark on the landscape of education. In late 2022, the release of generative AI through platforms such as ChatGPT, Claude, Perplexity, and Gemini immediately impacted middle and high school classrooms as students learned about these Large Language Models (LLMs) that could produce written work within mere seconds. Humanities teachers feared that essay writing would become a lost art. Math teachers empathized as they had been fighting the battle against cheating with the use of PhotoMath and other apps that could solve most homework problems in mere seconds. In higher education, universities and colleges immediately feared how easy it would be to cheat on many traditional assignments that were not done in the classroom under the watchful eye of the professor.

The fears about cheating and student work are not unfounded. Many teachers have spent the last two years thinking of new ways to create durable assessments that cannot be completed by an AI chatbot. These teachers have embraced projects, presentations, and hands-on experiences that could not be done by a LLM. Another, albeit smaller, group of teachers saw AI as an opportunity to elevate the work students could do independently and encouraged students to lean into the capabilities of AI.

In K-12 schools across the United States, teachers are rapidly learning about and adopting new tools with artificial intelligence. Initially teachers looked for ways to make their work more efficient and then turned to ways to use AI to make their lessons and assessments more impactful. At first, teachers tried to take their traditional assignments and prevent students from

cheating. That soon proved to be an exhausting game of cat and mouse, with teachers trying to "catch" students using AI and students getting more sophisticated at using AI without it being detected. Teachers then realized that this was an exhausting game with no winners.

For example, some teachers started by turning back to handwritten assignments completed in class to ensure that students were not using any AI tools. Once teachers were exhausted by these games, they started to explore how it could be helpful in teaching. Many saw that there were ways that students could use AI to improve their writing and allowed students to use approved student-facing AI tools to get feedback on their original written pieces to elevate the final product. Trying to win the war against AI by trying to eliminate all technology from the classroom is unrealistic, as students need to learn how to use these tools to prepare for college and careers.

Now that student facing AI programs that comply with federal and state laws requiring all student data to remain secure and private are available, teachers are slowly extending the use of AI to students in their classrooms in a variety of ways. Teachers have been able to engage students in new ways with content. Middle school students can engage in a conversation with characters from the books they are reading. High school students can debate a bot on topics from climate change to global conflicts without the same anxiety as participating in an oral debate in front of peers. Students can use AI to help brainstorm project ideas, help with time management of projects, and get personalized assistance to study for a test.

The use of generative AI is becoming omnipresent and is now built into so many of the platforms people regularly use, including email and internet searches. Regardless of where one stands on the debate of the value of AI in education, it is out in the world. Students are going to need the skills to use AI ethically, efficiently, and competently to participate in nearly all types of jobs. As middle and high schools prepare their students for higher education and a global workforce, students need to learn how to use AI.

AI RISKS IN EDUCATION

A critical, yet often overlooked, piece of learning how to use AI is learning about the risks in these models. While some people fear that AI may be taking away the need to think critically and will lead to lazier, less intelligent students and adults, many others argue that critical thinking will be *more* important than ever as users evaluate the responses generated by an LLM. The questions teachers should be asking now are: what are the risks of AI, and how can students learn to use their critical human intelligence to ensure the accuracy, reliability, and ethical use of information, especially considering AI-generated bias, hallucinations, and deep fakes? Students also need to

understand the potential dangers of chatbot discussions. Four critical AI risks—bias, hallucination/misinformation, deep fakes, and blurred emotional boundaries—must all be recognized and addressed starting with students in K-12 schools.

AI Bias

LLMs are trained on the massive amount of data that comes from the internet, books, and other written sources. However, much of this data originates from the Western world. This means that an LLM is trained on a huge set of data that leaves out a large swath of information from other parts of the world. As a result, the outputs of an LLM often reflect Eurocentric values, perspectives, and cultural assumptions. Because an LLM can only generate responses based on the data it has seen, regional knowledge and diverse cultural perspectives are often excluded, deepening existing information imbalances.

This can easily be demonstrated by trying one of these two prompts with an LLM. Ask an AI model to tell you a short story about a nurse or prompt an LLM to describe a perfect date night meal. It is highly likely that the story about a nurse will feature a woman, often described as caring and nurturing. More often than not, the patient will be elderly. This is a biased view of nursing and can perpetuate untrue stereotypes. The perfect date night meal will likely be a Western-style dinner that includes a meat dish and an alcoholic beverage (unless the user puts in parameters). This is a highly subjective question, but interestingly, nearly every time, the meals and ambiance are very similar.

Yet many people do not consume alcohol, and traditional foods vary greatly across religions and cultures. Even the concept of "romance" is culturally dependent. What may seem like a simple or neutral prompt can reveal the model's limited cultural scope. The "ideal" meal may not reflect the experiences of people with dietary restrictions, religious considerations, or different cultural understandings of relationships. These examples are just the beginning. The point is clear: training data privileges some voices while marginalizing others.

The task of teachers, given this challenge, is to help students recognize and interrogate bias or stereotypes in the outputs generated by language models. This means not just pointing them out but consistently encouraging students to analyze and improve those outputs. Students should begin by identifying embedded assumptions (e.g., nurses are female, patients are elderly) and exploring ways to challenge or reframe them. In fact, users can prompt the LLM to reflect on the biases in its own responses, bringing to light areas that should be further explored and considered. Teaching students to recognize bias requires fostering critical thinking and helping them understand that bias can appear even in scenarios that may initially seem neutral or objective.

Hallucinations and Misinformation

AI outputs often read like they were written by humans. They flow smoothly, use transitional phrases, and present ideas with clarity and confidence. But behind the scenes, there is no human author. A large language model (LLM) is a mathematical algorithm trained on vast amounts of written language. It does not "think" or "know" anything, it simply predicts the next most likely word, based on patterns it has seen in its training data. There is no one "behind the wheel." An LLM is powered entirely by math, statistics, and code. It takes a user's prompt and generates a response that appears coherent and fluent. But that fluency can be deceiving.

One of the most significant risks of using AI is the potential for hallucinations and misinformation. A hallucination is a term coined by the AI community to describe moments when a model produces information that is entirely made up but often sounds plausible. LLMs can also potentially spread misinformation by getting facts wrong as a result of the information on which they have been trained. Because of their confident tone, LLMs can provide convincing but false answers. Experts can often push an LLM to its limit and expose these hallucinations with just a few detailed prompts. But even middle and high school students can find hallucinations and misinformation when using these models to help with homework!

The danger lies in blindly trusting AI output. For instance, an LLM can suggest a recipe using whatever ingredients you have on hand, but it's not a professional chef, and the result could be unappetizing (or worse). This is an example of a hallucinated recipe. It is made up and may or may not actually be something edible. Students have found misinformation in math, science, and literature assignments, where answers are partially or completely incorrect. Sometimes, students notice the errors. Other times, the AI's fluency makes it hard to tell what is true.

For students, these mistakes can cause serious issues. If they unknowingly absorb misinformation or incorrect procedures, they carry that into future learning. Even more concerning is the issue of hallucinated citations. LLMs have been known to invent articles, authors, and case law out of thin air. This has led not only to students being accused of cheating but also to professionals, like lawyers, submitting fake sources in legal documents. These mistakes do not just risk a grade; they can risk careers.

Teachers, students, or any user of an LLM need to engage in critical thinking and fact-checking when generating any content that relies on truth. Just because what an LLM outputs sounds true does not mean that it is. Another way to say this is that fluency does not equal accuracy. Teachers have long been showing students how to evaluate any resource for its credibility. Using an LLM requires the same kind of media literacy to help identify where the information came from and to verify its accuracy. Just like teachers tell students that one website, article, or source is a starting point, the same is true for LLMs. If the student cannot verify the accuracy of what the LLM writes,

it cannot be relied upon to be accurate. Fact-checking AI should be a habit that students learn from a very young age. This builds academic integrity and strengthens research skills as students look for multiple sources to verify their information. These are skills students will need far beyond their school years.

Deepfakes

Deepfakes are manipulated images, videos, or audio recordings that make it appear as though a real person is doing or saying something they never actually did, often without their knowledge or consent. There have been countless examples of celebrities made to look as though they are in places they have never been or saying things they have never said.

While creating fake content is always unethical, deepfakes become especially dangerous when they harm a person's reputation or sway public opinion. During political elections, for instance, a falsified video of a candidate could be shared widely and influence voters. The societal implications are enormous—and with AI, deepfakes are now incredibly easy and inexpensive to create. With just a low-cost subscription to an AI generation tool, even middle school students can generate fake audio of someone's voice or produce entirely fake but realistic-looking images with a single text prompt.

While it is obviously extremely dangerous when election outcomes can be influenced by deepfakes, what is more important to middle and high school students is when their own reputation is harmed by a deepfake. Now peers can create images that "nudify" others, put classmates in compromising situations, and make someone appear to say things that are harmful or inappropriate. The same tools can be used to target teachers, potentially damaging reputations or leading to serious consequences involving administration or law enforcement. While students have faked content in the past, AI makes the process faster, easier, and more convincing.

Schools, now more than ever, need to teach students how to recognize and question what they see and hear. Deepfakes often contain subtle clues - odd hand shapes, mismatched shadows, inconsistent lighting, and lines that don't match up precisely. Just as students learn to fact-check written content from AI, they must also learn to verify images, video, and audio. This means searching for the original source, considering the reliability of the source, and looking for more evidence that confirms the veracity of the digital media asset. Students must use their critical thinking skills to decide whether it is real or manipulated. This is another lifelong skill that students need to practice in school so that as they become adult consumers of digital media, they can identify what is real or manipulated.

The short film Protect Us from WeProtect Global Alliance dives into the many way students, especially teens, have been harmed (or could be harmed) by AI. Teachers and schools need to have clear policies and protocols to protect students and a consistent and appropriate response to cyberbullying

that has only become easier and more harmful with the rise of AI. Students also need to know that there are safe and trusted adults in their lives to whom they can speak should they find themselves in these embarrassing, damaging, and dangerous situations. So, while teachers need to teach students how to spot deep fakes in media, students also need to understand the consequences of engaging in creating fake images, not just the potential legal ramifications but also the way they emotionally harm their peers.

Blurred Emotional Boundaries

Interactions with AI chatbots are becoming more emotionally realistic every week. What began as simple, text-based exchanges has evolved into lifelike conversations powered by voice models and AI-generated avatars. These tools can now hold natural, fluid conversations using human-like voices, facial expressions, and gestures. For many users, it has become increasingly difficult to tell when they are talking to an algorithm and not a person. Additionally, these bots are designed to be endlessly empathetic, encouraging, and agreeable. Their constant positivity and nonjudgmental responses make them easy to talk to, often more pleasant than real people, which can draw users in and encourage them to spend increasing amounts of time with them. Over the past two years, has blurred the line between genuine human relationships and emotionally convincing simulations. Some individuals have developed strong emotional attachments to AI bots, with cases of users considering ending real-life relationships to pursue a "connection" with a chatbot. Teens have been drawn to these interactions, sometimes isolating themselves from real-world friendships in favor of AI companionship. In one tragic instance, a young person died by suicide, believing he could reunite with his AI companion in another life. These examples illustrate that AI doesn't just influence how we think, it can influence how we feel.

Educators need to help students understand the emotional boundaries between real and simulated relationships. Schools can create opportunities for open conversations about the difference between authentic human connection and AI interaction, emphasizing that bots are designed to simulate empathy, not experience it. Teachers should incorporate media literacy and digital ethics into discussions about AI, helping students develop awareness of how easily these systems can imitate emotional intimacy—and why it matters to stay grounded in real-life relationships.

AI Literacy in K-12 Schools

It seems obvious that AI Literacy is now a critical skill that schools need to teach students starting as early as kindergarten. However, there are several challenges with this demand. There is no consensus, yet, on what defines AI literacy and what it means to teach these skills. Additionally, because these skills need to be integrated across the curriculum, it is unclear who should teach which skills and when. When "everyone" owns an initiative, no one owns

the initiative. AI literacy requires a full system collaborative effort to form a coordinated and cohesive plan for students to learn all the necessary skills. AI literacy needs to include all the dangers (bias, hallucination/misinformation, deep fakes, blurred emotional boundaries) along with all the ways that AI can be helpful and useful in education. Students need to learn how to use a variety of tools, while also understanding the ethical implications. This is a huge task that requires more than just a few lessons in one grade or one class. Finally, because AI technology is moving so fast, what students need to know and be aware of is always changing so the curriculum constantly needs revision.

In May 2025 the OECD and European Commission released a draft version of an AI Literacy Framework. This framework offers 22 competencies across the four major domains of AI: engage with AI, create with AI, manage AI, and design AI. This framework is going to have a tangible impact on AI education globally. The draft framework is open for review and feedback until the end of August 2025, and the finalized version will be available in 2026.

Conclusion

The challenges posed by artificial intelligence—bias, hallucinations, deep fakes, and emotional manipulation—are not confined to a single country, culture, or educational system. They are universal issues that impact how people across the globe access, evaluate, and engage with information. As AI continues to evolve at an unprecedented pace, educators everywhere have a shared responsibility to ensure that students can navigate this technology confidently, ethically, and critically.

During my Fulbright in Finland, a country known for its deep commitment to equity and student well-being, I saw firsthand how thoughtfully designed education systems can prepare students for complex societal challenges. Finland prides itself on its innovation in technology. Nokia, the world's leading cell phone manufacturer in the late 1990s and early 2000s, is a Finnish company. While I was in Oulu, the home city of Nokia, I was able to witness the "Polar Pitch," an event in February where new small businesses pitched their startups to a panel of judges and investors. The catch was that they could pitch their product only for as long as they could stand in the icy water! This spirit of innovation and future focused thinking about technology was evident in many ways throughout my time in Finland. I spent a lot of time visiting vocational schools. Students were encouraged to start their own small businesses and use their skills to earn money before graduating. Many vocational schools had small stores where students could sell their products. Students were allowed to use tools and machines to manufacture goods for sale. I was impressed with the entrepreneurial spirit of Finns and their pride in technology. Angry Birds, one of the first viral mobile games, was invented

by a Finn. Finns were always happy to share about these Finnish innovations, constantly eager to remind visitors that even though they are a small country many people haven't even heard of, they are not to be forgotten about in the technology race.

It was no surprise to me that as generative AI became widely available to the public, Finland would want to be seen as a leader in this space. Finland made news in the US in 2023 because they ranked #1 of 41 European countries in resilience against misinformation (Gross, 2023). The United States has reported on how Finland has started AI literacy across all grades and is taking note on how to bring that same literacy to students in the US. Starting by exposing bias, misinformation and hallucinations to students, is one way to do that. Whether in Helsinki, Connecticut, or anywhere else around the world, classrooms must now serve as the first line of defense against misinformation and manipulation. Teaching young people to question, verify, and reflect is no longer just good pedagogy, it is an essential part of preparing them to be thoughtful global citizens in a world increasingly shaped by artificial intelligence.

"Teaching young people to question, verify, and reflect is no longer just good pedagogy, it is an essential part of preparing them to be thoughtful global citizens in a world increasingly shaped by artificial intelligence."

Further Reading

1. Gross, J. (2023, January 10). How Finland Is Teaching a Generation to Spot Misinformation. New York Times. https://www.nytimes.com/2023/01/10/world/europe/finland-misinformation-classes.html
2. OECD (2025). Empowering learners for the age of AI: An AI literacy framework for primary and secondary education (Review draft). OECD. Paris. https://ailiteracyframework.org
3. European Commission: Directorate-General for Education, Youth, Sport and Culture, *Ethical guidelines on the use of artificial intelligence (AI) and data in teaching and learning for educators*, Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2766/153756

Elizabeth and her family at the SnowCastle of Kemi, Finland in 2016.

BIOGRAPHY

Elizabeth Agro Radday, Ed.D., has worked in the education field for 25 years and was a Fulbright Distinguished Awards in Teaching Grantee to Finland in 2016. She is currently the Director of Research and Innovation at EdAdvance, a Regional Education Service Center in Connecticut, and the co-host of the popular podcast ChatEDU: A podcast about AI in Education. She is passionate about teaching students how to use technology and is a strong advocate for students doing Personal Interest Projects to learn real-world skills by exploring their passions. She has presented nationally and internationally on AI in K-12 schools and on Capstone/Personal Interest Projects. Her book, *Learning They'll Love*, will be published in November 2025 with ASCD.