Responsible Use of Artificial Intelligence

Date: 2024
From: Gale In Context Online Collection
Publisher: Gale, part of Cengage Group
Document Type: Topic overview
Length: 1,218 words
Content Level: (Level 5)
Lexile Measure: 1310L

Full Text:
Artificial intelligence (AI) is a technology that allows computers to perform tasks that historically have required human intelligence, such as problem solving and making decisions. AI has almost limitless potential in the future. However, there are many aspects of AI that can lead to inappropriate, unethical, or even unlawful use. As AI tools become more advanced and more widely available, developers, students, workers, and others should familiarize themselves with the concept of “responsible AI.” They should follow certain guidelines that can reduce the chances of AI being used in a way that is harmful to oneself or others.

Critical Thinking Questions

- Why do students need to know about responsible AI?
- What are some ways that companies can create more equitable data for AI applications?
- Why is transparency an important aspect of responsible AI from both the developer and user standpoint?

What is AI?

AI is a machine-based system designed by humans to make recommendations, decisions, and predictions in order to simulate intelligent behavior. There are different kinds of processes and functions that can go into an AI application, and many applications combine these functions to create a seamless user experience. One of these functions is machine learning, where machines “learn” how to do a task based on feedback from data, humans, or algorithms. Another is natural language processing, which helps machines interpret different human languages. Some examples of AI found in everyday life are digital assistants, navigation apps, autcomplete in emails, autocorrect in text messages, and more.

Responsible AI and Developers

AI can be helpful in many different applications, but because AI is creator-driven and can be applied to society on a broad scale, biases, assumptions, and mistakes made by the creators can be amplified and have a huge effect on the world. For example, AI can be helpful in reviewing applications for bank loans or predicting what illness a patient might have based on the data collected at a doctor’s visit. However, when incorrect data or biases like racism or sexism are unconsciously added to the machine learning model, there can be negative effects. For example, racial minorities may be unfairly denied loans or female patients may be misdiagnosed.

AI applications like bots have been used for political reasons to spread misinformation and influence people to support candidates or causes. Other issues that developers should consider include lack of privacy, potential theft of user information, and the environmental cost of running AI systems. A great deal of computing power and electricity is needed to train and run AI models, resulting in high carbon emissions that can harm the environment.

To combat these problems, experts at the National Institute of Standards and Technology (NIST) created a list of seven principles to guide companies as they develop AI applications. Some of these principles revolve around making the AI system transparent and accountable, which would increase trust that the system is functioning as it should. Another principle is to make the AI interpretable by the user, so that they can understand how the AI came to its output. There is a principle associated with mitigating potentially harmful bias by eliminating discrimination and providing equitable and fair data and output. This is perhaps the most challenging principle, so experts recommend that companies get external analysts to review their application for unconscious biases and equity. The other principles focus on making the AI application secure against attacks from hackers, safe for people to use and not endanger the environment, and ensure user privacy.

Responsible AI and Users

Users such as students and workers must also be careful when using AI platforms for various purposes. ChatGPT and other AI applications have been known to produce misinformation, called hallucinations. This can lead to everything from a bad grade on a history paper to lawsuits. For example, in 2023, OpenAI, the company that created ChatGPT, was sued for defamation after the model created text that falsely accused a Georgia radio host of embezzlement. ChatGPT generated this text for a user who was researching a real legal case.

The legal questions around AI and copyright and intellectual property have not yet been decided. For example, if AI companies use copyrighted material to train their models, and that material is used to generate new content, it may count as plagiarism and trigger lawsuits and action against users and AI companies. US law does now allow copyright protection for content that is solely AI-generated. In addition, using AI-generated content without disclosing that the content is not original work may be considered plagiarism.

Privacy and protection of personal information is another important issue for users. Many AI models retain and analyze the prompts people are putting into the AI as a way to continue their "learning." Any personal information that a user enters into the prompt is at risk of being taken in a hacker attack.

Students, workers, and others who use content from AI applications need to critically evaluate the AI output themselves, and ensure for themselves that the output from the AI is accurate, unbiased, and free for them to use. Some questions user may want to ask themselves are:

- Can I find credible sources that support the data that the AI application shared?
- Can I test the accuracy of the output of the AI application, and do the results match up with what the application generated?
- Is the data inclusive of factors such as gender, race, and marginalized groups that might have been overlooked?
- Am I sharing or asked to share personal information with the AI application that I am not comfortable sharing or that would put me at risk?
- Knowing that application might be collecting the data that I share in my prompt, do I need to change the prompt to make sure my information stays secure?

School Policies on AI

Some teachers and schools prohibit the use of AI entirely, so students should be aware of what the individual policies are that may affect them. Other teachers and schools have opted instead to train students to use AI responsibly. This might include training students to use critical thinking to examine the output of AI applications, teaching students about how racial, gender, and other biases can be unwittingly built into AI applications, and modeling transparency about AI use and modeling appropriate AI practices. Some teachers conduct AI practice with their students to show real-world applications and the results of using AI for particular tasks.

Regardless of whether AI-generated content is allowed for classroom use or not, teachers and administrators must produce clear guidelines for AI use, including what constitutes plagiarism. If those guidelines are not clear, students should approach teachers with their questions to make sure they understand what the expectations are for AI use. Additionally, students should be transparent about the use of AI in their work, and follow regulations and laws around AI to protect themselves and others.

Words to Know

AI model

A program that uses a group of data sets to find, analyze, and synthesize patterns.

developer

Someone who creates new software products, often using coding.

intelligent behavior

When a machine identifies patterns, reasons, acts, and learns in a way that parallels human cognition.

machine learning

The creation of computer systems that can retain and synthesize information without being explicitly directed to do so by a developer or user.

natural language processing

A branch of computer science and AI that analyzes text and human speech so computers can understand and interpret them.

Source Citation (MLA 9th Edition)

"Responsible Use of Artificial Intelligence." *Gale In Context Online Collection*, Gale, 2024. *Gale In Context: Middle School*, link.gale.com/apps/doc/YFBVOK941237208/MSIC?u=nysl_ro_cmhs&sid=bookmark-MSIC&xid=6deab9f5. Accessed 8 Jan. 2025.

Gale Document Number: GALE|YFBVOK941237208