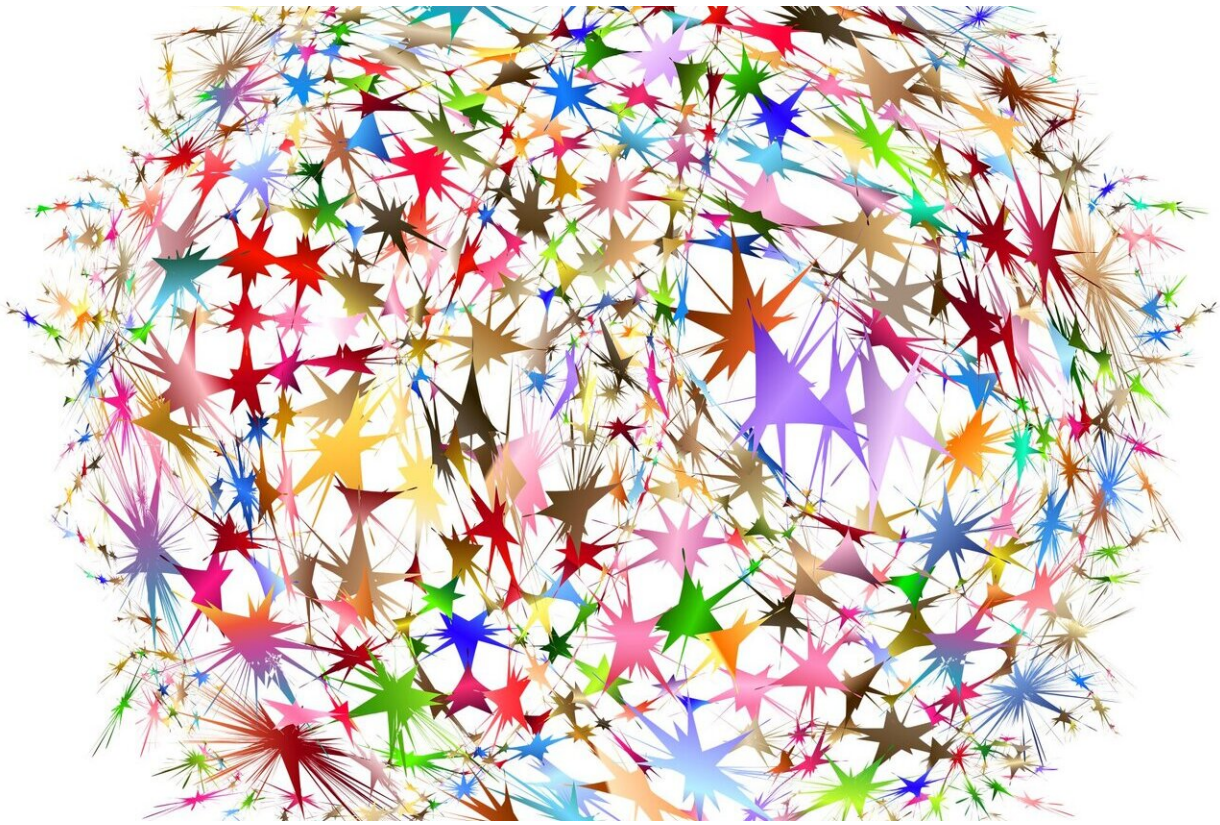
Forget dystopian scenarios—AI is pervasive today, and the risks are often hidden

November 22 2023, by Anjana Susarla



Credit: CC0 Public Domain

The turmoil at ChatGPT-maker OpenAI, sparked by the board of directors firing high-profile CEO Sam Altman on Nov. 17, 2023, has put a spotlight on artificial intelligence safety and concerns about the rapid

development of artificial general intelligence, or AGI. AGI is loosely defined as human-level intelligence across a range of tasks.

The OpenAI board stated that Altman's termination was for lack of candor, but speculation has centered on a rift between Altman and members of the board over concerns that OpenAI's remarkable growth—products such as ChatGPT and Dall-E have acquired hundreds of millions of users worldwide—has hindered the company's ability to focus on catastrophic risks posed by AGI.

OpenAI's goal of developing AGI has become entwined with the idea of AI acquiring superintelligent capabilities and the need to safeguard against the technology being misused or going rogue. But for now, AGI and its attendant risks are speculative. Task-specific forms of AI, meanwhile, are very real, have become widespread and often fly under the radar.

As a researcher of information systems and responsible AI, I study how these everyday algorithms work—and how they can harm people.

AI is pervasive

AI plays a visible part in many people's daily lives, from face recognition unlocking your phone to speech recognition powering your digital assistant. It also plays roles you might be vaguely aware of—for example, shaping your social media and online shopping sessions, guiding your video-watching choices and matching you with a driver in a ride-sharing service.

AI also affects your life in ways that might completely escape your notice. If you're applying for a job, many employers use AI in the hiring process. Your bosses might be using it to identify employees who are likely to quit. If you're applying for a loan, odds are your bank is using

AI to decide whether to grant it. If you're being treated for a medical condition, your health care providers might use it to assess your medical images. And if you know someone caught up in the criminal justice system, AI could well play a role in determining the course of their life.

Algorithmic harms

Many of the AI systems that fly under the radar have biases that can cause harm. For example, machine learning methods use inductive logic, which starts with a set of premises to generalize patterns from training data. A machine learning-based resume screening tool was found to be biased against women because the training data reflected past practices when most resumes were submitted by men.

The use of predictive methods in areas ranging from health care to child welfare could exhibit biases such as cohort bias that lead to unequal risk assessments across different groups in society. Even when legal practices prohibit discrimination based on attributes such as race and gender—for example, in consumer lending—proxy discrimination can still occur. This happens when algorithmic decision-making models do not use characteristics that are legally protected, such as race, and instead use characteristics that are highly correlated or connected with the legally protected characteristic, like neighborhood. Studies have found that risk-equivalent Black and Latino borrowers pay significantly higher interest rates on government-sponsored enterprise securitized and Federal Housing Authority-insured loans than white borrowers.

Another form of bias occurs when decision-makers use an algorithm differently from how the algorithm's designers intended. In a well-known example, a neural network learned to associate asthma with a lower risk of death from pneumonia. This was because asthmatics with pneumonia are traditionally given more aggressive treatment that lowers their mortality risk compared to the overall population. However, if the

outcome from such a neural network is used in hospital bed allocation, then those with asthma and admitted with pneumonia would be dangerously deprioritized.

Biases from algorithms can also result from complex societal feedback loops. For example, when predicting recidivism, authorities attempt to predict which people convicted of crimes are likely to commit crimes again. But the data used to train predictive algorithms is actually about who is likely to get re-arrested.

AI safety in the here and now

The Biden administration's recent executive order and enforcement efforts by federal agencies such as the Federal Trade Commission are the first steps in recognizing and safeguarding against algorithmic harms.

And though large language models, such as GPT-3 that powers ChatGPT, and multimodal large language models, such as GPT-4, are steps on the road toward artificial general intelligence, they are also algorithms people are increasingly using in school, work and daily life. It's important to consider the biases that result from the widespread use of large language models.

For example, these models could exhibit biases resulting from negative stereotyping involving gender, race or religion, as well as biases in the representation of minorities and disabled people. As these models demonstrate the ability to outperform humans on tests such as the bar exam, I believe that they require greater scrutiny to ensure that AI-augmented work conforms to standards of transparency, accuracy and source crediting, and that stakeholders have the authority to enforce such standards.

Ultimately, who wins and loses from large-scale deployment of AI may

not be about rogue superintelligence but about understanding who is vulnerable when algorithmic decision-making is ubiquitous.

Provided by The Conversation

Citation: Forget dystopian scenarios—AI is pervasive today, and the risks are often hidden (2023, November 22) retrieved 9 May 2024 from https://techxplore.com/news/2023-11-dystopian-scenariosai-pervasive-today-hidden.html