

## CLASS ACT

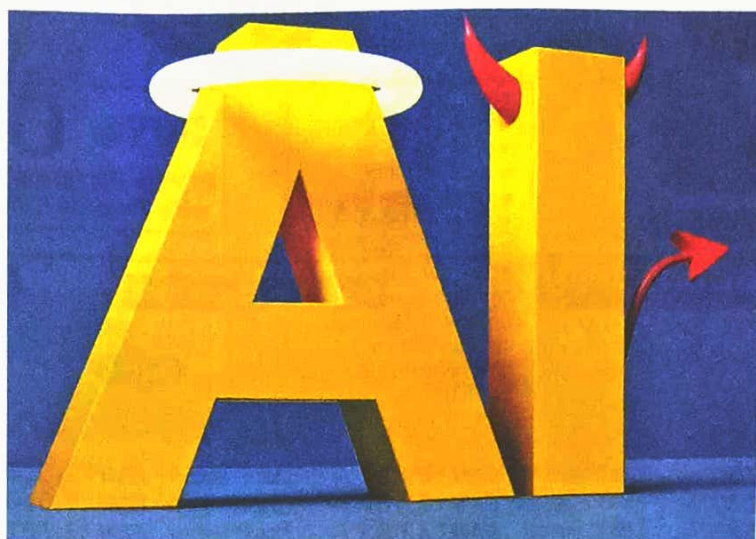
# Learning about the machine

Things about AI the youth should know. **Debabratee Dhar** asks an expert

**A**t IIT-Kharagpur, they teach a course called AI and Ethics. The instructor is Animesh Mukherjee whose research has tackled some of the pressing challenges in AI ethics, from bias mitigation in machine learning models to the safety and reliability of AI-driven platforms. The reason for teaching such a course to would-be-engineers is this, according to Mukherjee. "Using AI is the need of the hour. Previously, AI was confined to the realms of research. Now that AI has become a commodity and is widely used, there could be good and inadvertent consequences," he says. As per the course outline, there are no books, though there is a reading list in case students want to refer to it. Should AI and ethics be taught in schools as well? Says Mukherjee, "If it is, it should be structured, practical and intellectually honest about trade-offs. It should address questions like, 'Should I trust AI to help with homework?' or 'Is it okay to use AI in exams?'" Here, Mukherjee answers questions about issues he feels the youth should be aware of. Excerpts:

**Q Can AI systems be made free from bias in the real world? I am asking about prompts of e-commerce websites or search engine results.**

We can see inadvertent consequences across platforms, be it social media or e-commerce. That's because we do not know what kind of AI algorithm they operate under. How the various shopping platforms prompt items to buyers, some brands over others, it's all very opaque. Such algorithms that we cannot "see" (do not have access to) are called black box algorithms. The algorithms that we can "see" are called white box, and it is possible to develop a fair algorithm for them. And many researchers and practitioners are developing



ISTOCK.COM/CAGKANSAYIN

such algorithms.

**Q What about the facial recognition bias in AI systems based on Western datasets?**

In law enforcement, facial recognition is a very common tool. Surveillance cameras are used for capturing these images. In case of a false positive, there is a massive social repercussion. We read about this case where an elderly gentleman went to a shopping mall and came back home without buying anything. He was captured in the mall's CCTV footage. Later, in connection with a case of shoplifting, the police came to his house and accused him. It was done on the basis of face-matching through AI datasets. The police got that data from the CCTV footage, and based on that, he was arrested. And then this man had to go through a trial to prove his innocence. In the West, these models are trained on their own population and yet, such biases exist. Now, if we used these models in our country, where ethnicity, facial features,

complexion, everything differs, there would be even more cases of such false matches.

**Q There is also some talk of AI-governance.**

Governance is done through laid-out rules and policies. These policies have to be fed into the computer. When policies are operationalised, there are several interpretations, and AI will behave differently for each policy. One cannot possibly feed all permutations and combinations of interpretations to the AI. Besides, AI only understands algorithms, while an interpretation requires a more human approach, which also requires philosophising. Accountability becomes a big issue. For example, if an automatic car meets with an accident, who is to blame — the traffic controller, the software engineer who programmed the car or AI itself?

It is also difficult to obtain good, unbiased data. Training AI systems with noisy and biased data might be very problematic. If all the interpreta-

tions of the policies that we are feeding into AI are made public, some of this dilemma can be mitigated. Any good AI software should release a data card and a model card. The model card tells us about the parameters, all the interpretations used and even the energy consumed to train the AI model. The data card precisely describes the data, its source, steps taken to clean it and the annotation procedure.

**Q How can AI be used for real-time crisis response in rural India?**

Disaster management works through several disaster response systems. We have many volunteers. They collect data on how much food or medicines need to be sent and where. When a disaster strikes, other than rescue workers, local people also post about the situation. At IIT-Kharagpur, we have the DisARM project. Basically, we collect social media data and sift through situational and non-situational information. Helpline numbers, how much food and medicines are needed, how many people are stranded — these are situational information. Fake and irrelevant news qualify as non-situational information.

**Q This time AI was used in the West Bengal Assembly elections to help parties get a sense of who the Bengali is. Can AI systems be tuned to understand nuanced linguistic or cultural diversity?**

It's very difficult. But people these days are trying to make multi-objective and pro-social systems, meaning systems that can handle cultural nuances, linguistic and contextual differences.

**Q In these AI times will students still care for classroom teaching?**

Students don't skip class just because AI is there. If teaching becomes routine and mundane, then students will lose interest. There is no alternative to a good teacher.