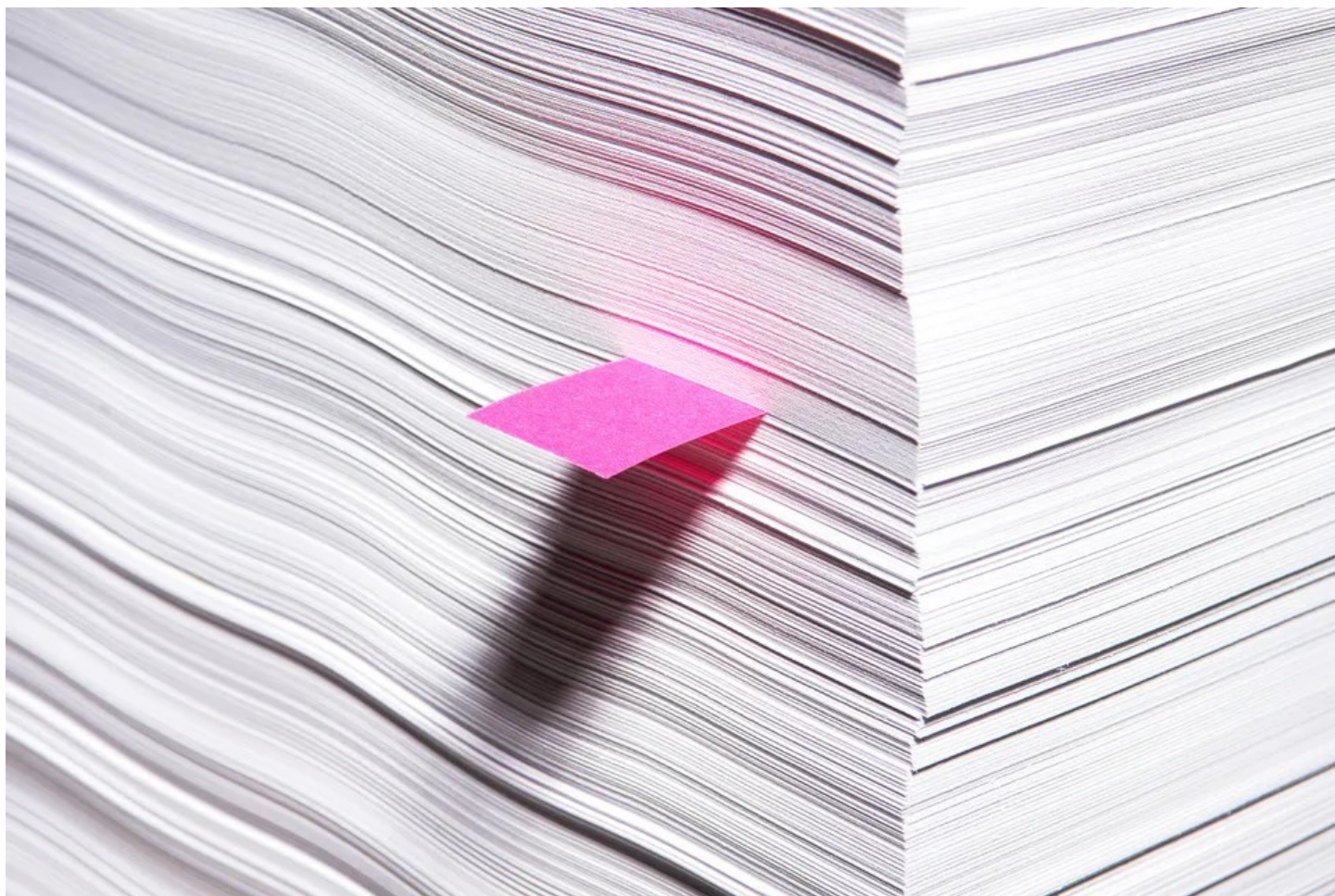WILL KNIGHT   BUSINESS   AUG 10, 2022 7:00 AM

Sloppy Use of Machine Learning Is Causing a 'Reproducibility Crisis' in Science

AI hype has researchers in fields from medicine to sociology rushing to use techniques that they don't always understand—causing a wave of spurious results.

PHOTOGRAPH: PM IMAGES/GETTY IMAGES

🗝 The AI Database →

APPLICATION: PREDICTION   END USER: RESEARCH

SECTOR: HEALTH CARE, PUBLIC SAFETY, RESEARCH   TECHNOLOGY: MACHINE LEARNING

HISTORY SHOWS CIVIL wars to be among the messiest, most horrifying of human affairs. So Princeton professor Arvind Narayanan and his PhD student Sayash Kapoor got suspicious last year when they discovered a strand of political science research claiming to predict when a civil war will break out with more than 90 percent accuracy, thanks to artificial intelligence.

A series of papers described astonishing results from using machine learning, the technique beloved by tech giants that underpins modern AI. Applying it to data such as a country's gross domestic product and unemployment rate was said to beat more conventional statistical methods at predicting the outbreak of civil war by almost 20 percentage points.

Yet when the Princeton researchers looked more closely, many of the results turned out to be a mirage. Machine learning involves feeding an algorithm data from the past that tunes it to operate on future, unseen data. But in several papers, researchers failed to properly separate the pools of data used to train and test their code's performance, a mistake termed "data leakage" that results in a system being tested with data it has seen before, like a student taking a test after being provided the answers.

"They were claiming near-perfect accuracy, but we found that in each of these cases, there was an error in the machine-learning pipeline," says Kapoor. When he and Narayanan fixed those errors, in every instance they found that modern AI offered virtually no advantage.

That experience prompted the Princeton pair to investigate whether misapplication of machine learning was distorting results in other fields—and to conclude that incorrect use of the technique is a widespread problem in modern science.

AI has been underlined as potentially transformative for science because of its capacity to unearth patterns that may be hard to discern using more conventional data analysis. Researchers have used AI to make breakthroughs in predicting protein structures, controlling fusion reactors, probing the cosmos.

Yet Kapoor and Narayanan warn that AI's impact on scientific research has been less than stellar in many instances. When the pair surveyed areas of science where machine learning was applied, they found that other researchers had identified errors in 329 studies that relied on machine learning, across a range of fields.

Kapoor says that many researchers are rushing to use machine learning without a comprehensive understanding of its techniques and their limitations. Dabbling with the technology has become much easier, in part because the tech industry has rushed to offer AI tools and tutorials designed to lure newcomers, often with the goal of promoting cloud platforms and services. "The idea that you can take a four-hour-long online course and then use machine learning in your scientific research has become so overblown," Kapoor says. "People have not stopped to think about where things can potentially go wrong."

Excitement around AI's potential has prompted some scientists to bet heavily on its use in research. Tonio Buonassisi, a professor at MIT who researches novel solar cells, uses AI extensively to explore novel materials. He says that while it is easy to make mistakes, machine learning is a powerful tool that should not be abandoned. Errors can often be ironed out, he says, if scientists from different fields develop and share best practices. "You don't need to be a card-carrying machine-learning expert to do these things right," he says.

Kapoor and Narayanan organized a workshop late last month to draw attention to what they call a "reproducibility crisis" in science that makes use of machine learning. They were hoping for 30 or so attendees but received registrations from over 1,500 people, a surprise that they say suggests issues with machine learning in science are widespread.

During the event, invited speakers recounted numerous examples of situations

where AI had been misused, from fields including medicine and social science. Michael Roberts, a senior research associate at Cambridge University, discussed problems with dozens of papers claiming to use machine learning to fight Covid-19, including cases where data was skewed because it came from a variety of different imaging machines. Jessica Hullman, an associate professor at Northwestern University, compared problems with studies using machine learning to the phenomenon of major results in psychology proving impossible to replicate. In both cases, Hullman says, researchers are prone to using too little data, and misreading the statistical significance of results.

Keep Reading
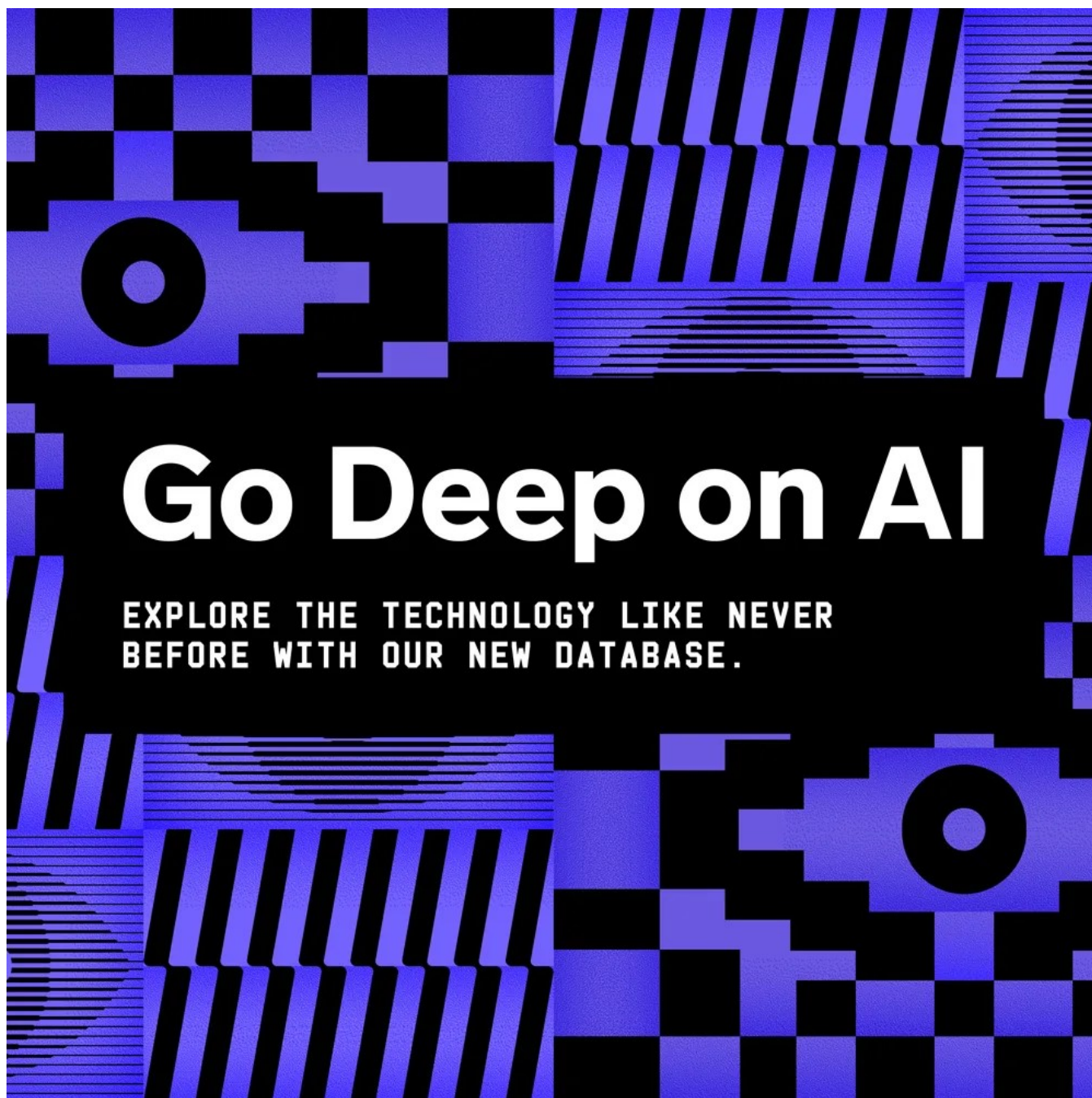
Search our artificial intelligence database and discover stories by sector, tech, company, and more.

Momin Malik, a data scientist at the Mayo Clinic, was invited to speak about his own work tracking down problematic uses of machine learning in science. Besides common errors in implementation of the technique, he says, researchers sometimes apply machine learning when it is the wrong tool for the job.

Malik points to a prominent example of machine learning producing misleading results: Google Flu Trends, a tool developed by the search company in 2008 that

aimed to use machine learning to identify flu outbreaks more quickly from logs of search queries typed by web users. Google won positive publicity for the project, but it failed spectacularly to predict the course of the 2013 flu season. An independent study would later conclude that the model had latched onto seasonal terms that have nothing to do with the prevalence of influenza. "You couldn't just throw it all into a big machine-learning model and see what comes out," Malik says.

Some workshop attendees say it may not be possible for all scientists to become masters in machine learning, especially given the complexity of some of the issues highlighted. Amy Winecoff, a data scientist at Princeton's Center for Information Technology Policy, says that while it is important for scientists to learn good software engineering principles, master statistical techniques, and put time into maintaining data sets, this shouldn't come at the expense of domain knowledge. "We do not, for example, want schizophrenia researchers knowing a lot about software engineering," she says, but little about the causes of the disorder. Winecoff suggests more collaboration between scientists and computer scientists could help strike the right balance.

See What's Next in Tech With the Fast Forward Newsletter

From artificial intelligence and self-driving cars to transformed cities and new startups, sign up for the latest news.

Your email

Enter your email

By signing up you agree to our User Agreement (including the class action waiver and arbitration provisions), our Privacy Policy & Cookie Statement and to receive marketing and account-related emails from WIRED. You can unsubscribe at any time.

While misuse of machine learning in science is a problem in itself, it can also be seen as an indicator that similar issues are likely common in corporate or government AI projects that are less open to outside scrutiny.

Malik says he is most worried about the prospect of misapplied AI algorithms causing real-world consequences, such as underlining denying someone medical care or unjustly advising against parole. "The general lesson is that it is not appropriate to approach everything with machine learning," he says. "Despite the rhetoric, the hype, the successes and hopes, it is a limited approach."

Kapoor of Princeton says it is vital that scientific communities start thinking about the issue. "Machine-learning-based science is still in its infancy," he says. "But this is urgent—it can have really harmful, long-term consequences."

More Great WIRED Stories

- 📩 The latest on tech, science, and more: Get our newsletters!
- The high-stakes race to engineer new psychedelic drugs
- The hacking of Starlink terminals has begun
- What to do with your old PlayStation 4
- Three possible futures of the monkeypox epidemic
- We interviewed Meta's AI chatbot about … itself
- ◉ Explore AI like never before with our new database
- 💻 Upgrade your work game with our Gear team's favorite laptops, keyboards, typing alternatives, and noise-canceling headphones

Will Knight is a senior writer for WIRED, covering artificial intelligence. He was previously a senior editor at *MIT Technology Review*, where he wrote about fundamental advances in AI and China's AI boom. Before that, he was an editor and writer at *New Scientist*. He studied anthropology and journalism in… Read more

SENIOR WRITER 🐦