"Neither fire nor ice - just chatter"

Haynes Miller

Once upon a time we worried that the computer would take over our spaceship, or our job. Now computers, armed with novel artificial intelligence algorithms, threaten apocalypse through something much more mundane: an unstoppable flood of alternative facts and false information.

ChatGPT and Bard and other chatbots have caused a panic in academia because of their uncanny ability to write convincing freshman essays. Mathematicians are not so impressed, because these apps are famously bad at arithmetic. This is a temporary glitch, which Steven Wolfram is hard at work to repair. But the deeper problem, again perhaps clearer in mathematics than in some other disciplines, is that if some bit of information can't be found on the web, they will just make things up. When challenged, they will apologize or make a joke.

This is the first problem: These applications are spectacular purveyors of false information. They never reveal sources; there are no footnotes. And after all these assertions are computer generated -- so they must be correct, right? because computers don't make mistakes. This flaw in the preservation of truth will grow exponentially as the chatbots increasingly prey on each other, magnifying their own false statements.

They also represent a direct threat to our system of democracy -- already under threat from human agents. Very soon when a federal agency opens a comment period, it will be flooded with highly believable comments from many apparently genuine concerned citizens and interest groups -- thousands of comments generated from prompts created by programmers working for wealthy individuals or wealthier corporations. The comment inboxes of our elected officials and our newspapers will fill up with opinions from myriad constituents, exhibiting a variety of styles and subtle variations of point of view -- and all fake.

US security services are rightly concerned about the potential for attacks by foes, as well as friendly competitors, creating panics or military errors. And they are hard at work weaponizing this technology for their own potential use. It's not too different from a biological weapon.

I am far from alone in fearing this future; an excellent article appeared already in The New York Times in January, and more recently Turing Award winner

Geoffrey Hinton has resigned from Google in order to warn of these risks.

Here are two suggestions for mitigating this looming disaster.
(1) Every fragment of AI generated text should bear an indelible watermark identifying it as such -- the way we mark cigarette packages as dangerous, or list ingredients on food packaging. This is a highly complex computer science challenge! -- one that our School of Computing, with its avowed ethical commitment, should throw itself into immediately and forcefully.
(2) Every statement of “fact” made by generative AI should be footnoted with a reference to its source. This could be done in a way that doesn’t interrupt the flow of the text, by hyperlinks or by a link to a separate page. (Thanks to my colleague Franz Ulm for this suggestion.)

This threat must be contained, by law and treaty, as biological weapons have been. Suggesting effective legislation is a major "design" project.
This could be a great collaboration, proving the value of the system of "bridge appointments" created when the College of Computing was founded: It will involve Political Science, Linguistics, Literature, Comparative Media Studies. This would be a wonderful issue on which MIT, in its current computational phase, could take the lead in dramatic style.

A good start, very helpful in moving a national discussion of this clear and present danger, would be a detailed and forceful statement of the dangers of this technology by the leadership of the College.