Check against delivery

Keynote panel remarks by Patricia Kosseim, Information and Privacy Commissioner of Ontario — IAPP Canada Privacy Symposium 2023
May 26, 2023

When I wrote my blog, *Privacy and Humanity on the Brink*, over a year ago, I compared the future of artificial intelligence (AI) to other existential threats like global warming, and potential nuclear war. I wrote about AI increasingly crossing the boundary between predicting human behaviour with near-perfect accuracy and nudging our behaviour in ways that jeopardize our sense of human agency. Or, as Daniel Solove has put it, "algorithms not only predict the future; they create it."

I wrote about the power of algorithms to influence the educational opportunities we offer (or not) to our children, the kinds of jobs we are likely to get or even apply for, how we make sentencing decisions and bail assessments, who we choose to marry or vote for, and what kinds of consequential health and life decisions we make based on our genomic makeup and statistical likelihoods.

I wrote about how social media platforms and personalized newsfeeds influence the kind of information we're exposed to and reinforce our already existing thoughts, predispositions, attitudes and biases based on what we read and write, what we like and who our friends are.

And that was before Chat GPT became widely available for mass consumption. Since its public release in March, the world consciousness awoke to the power of generative AI and large language models to create synthetic information — including disinformation — showing us concrete examples of how deep fakes can destabilize the truth and further distort our understanding of the world around us.

As regulators, we need to be wary about sounding alarmist, exaggerating risks or fear-mongering, lest we be seen as being out of touch with reality. We can't cry wolf without risk of losing our credibility, persuasiveness and relevance. So, is it an exaggeration to speak of the potential of AI to push privacy and humanity to the brink?

The now-familiar sci-fi plot of robots taking over the world can be traced back to 1921, in a stage play called *Rossum's Universal Robots*, by Czech playwright Karel Čapek, long before computers, robots or AI were even imaginable.

But there are also more recent warnings that are being sounded by world-leading scientific experts with a front row seat to the development of AI and its prospective risks.

In a BBC interview in 2014, Stephen Hawking warned against the potential for fully developed AI to take off on its own at an ever-increasing rate, that could "could spell the end of the human race."

Five years ago, the Pew Research Centre canvassed the views of almost one thousand leading technology experts, business leaders, policymakers, and activists on potential impacts of AI

systems and published their findings in a report called *Artificial Intelligence and the Future of Humans*. Most experts interviewed were deeply concerned about the long-term threats AI technologies pose to human autonomy, agency, and capabilities, and what they called the essential elements of being human.

Geoffrey Hinton, dubbed the godfather of AI, recently left his job at Google so he could be free to speak about the risks and benefits of AI systems. In his informed view, it is very reasonable to be worrying about these issues now. We're getting close to computers being able to improve themselves in a manner we can no longer control which could "mean the end of people." He worries about the impacts of artificial general intelligence on humanity, particularly when used by autocracies for malevolent purposes. He's calling for the need for international treaties to deal particularly with autonomous lethal weapons. Unless everyone is sensible about how they use AI technologies, he says, it is not inconceivable for AI to wipe out humanity.

Yoshua Bengio, Elon Musk, and other prominent technologists signed an open letter called *Pause Giant AI Experiments*. The letter calls for an immediate six-month moratorium on the development of AI systems more powerful than GPT-4 during which researchers should focus on making today's AI systems "more accurate, safe, interpretable, transparent, robust, aligned, trustworthy and loyal." The letter cites one of the Alisomar AI principles emphasizing that "advanced AI could represent a profound change in the history of life on Earth, and should be planned for and managed with commensurate care and resources."

Historian, philosopher and futurist, Yuval Noah Harari has described these large language models as having "hacked the operating system of human civilization" in a manner that actually threatens our survival. Language, he says, "is the stuff … human culture is made of." What will happen, he asks, when non-human intelligence becomes better than we are at generating stories and creating new cultural ideas and artifacts through "mass-produced political content, fake-news stories and scriptures for new cults." Or when they become increasingly good at mimicking sentient feelings and intimacy to manipulate us into changing our opinions and worldviews. Just as nuclear technology could be used to physically destroy human civilization, new AI models too have the potential to be used as a weapon of mass destruction of sorts that can destroy our social world. "What we are talking about," he says, "is potentially the end of human history. Not the end of history, just the end of its human-dominated part."

Technologies have brought humanity to the brink before, and yet we have deliberately chosen not to take the plunge. For example, we have decided as an international community to ban human cloning, and although we allow research of some human-animal chimeras to grow cells or tissues, never beyond a certain stage of development.

As the ethical adage goes, just because we could doesn't mean we should.

Just a couple of days ago, Sam Altman and his colleagues at Open AI, released a statement) describing superintelligent AI systems as being more powerful than any other technological invention we have dealt with before. "Given the possibility of existential risk, we can't just be reactive." Comparing superintelligence to nuclear energy, they call for the creation of something like an International Atomic Energy Agency to carry out inspections, audit for compliance with safety standards, and restrict deployment of capabilities above a certain threshold.

At the same time, they also seem to suggest that it's important to allow companies to pursue the development of AI systems below a significant capability threshold without burdensome regulation.

My own view is that we have to do *both*. We must act urgently, nationally and internationally, to prohibit dangerous deployment of AI systems and set clear boundaries beyond which as a global society we all agree we should not go.

But just as urgently, we must act at a local level to regulate the space of permissible AI technologies we want to allow to advance the public good, on certain terms and conditions and with appropriate oversight to ensure safety, fairness, transparency, accountability, and privacy. We need a robust set of rules to determine what are beneficial uses, for whom, by whom, and who gets to decide.

Yesterday, my office, together with the Ontario Human Rights Commission, issued a joint statement urging the Ontario government to develop and implement effective guardrails on the public sector's deployment of AI technologies to ensure that Ontario can reap the benefits of AI technologies, but in a manner that is ethically responsible, accountable, sustainable, and supported by public trust.

Although we commend the foundation proposed by government in its 2021 Trustworthy Artificial Intelligence (AI) Framework, and related draft principles and guidelines, it is urgent to press forward with that initiative and build on that momentum to establish a binding set of robust and granular rules for public sector use of AI technologies. Such rules are all the more necessary considering that the *Federal Artificial Intelligence Act*, whatever its fate, will not cover Ontario's public sector. The government must act now to fill that vacuum.

In sum, for data to be used for good, we have to stop data for bad. And for data to be used for good, we need *good* data.

We must take steps to ban dangerous and harmful uses of AI, including generative AI systems, that threaten not only our physical well-being as a species, but can, through the deliberate generation of disinformation, potentially undermine our social cohesion and unravel the very fabric that keeps us connected.

We need to get these malevolent uses of AI systems off the table, while also creating a healthy space for beneficial uses of AI systems to be developed and deployed within a robust regulatory governance framework. We need clear guardrails to ensure that fair datasets are used to break, rather than reinforce, the cycles of systemic historical biases that perpetuate social division and inequality. Only with effective guardrails and fair, accurate and legitimately sourced data, can institutions and organizations make use of data for good in a way that earns and maintains social trust, while simultaneously being used to address society's pressing challenges.