The risks of AI are real but manageable

The world has learned a lot about handling problems caused by breakthrough innovations.

By Bill Gates published on Tuesday, Jul 11, 2023
Tech thinking

The risks created by artificial intelligence can seem overwhelming. What happens to people who lose their jobs to an intelligent machine? Could AI affect the results of an election? What if a future AI decides it doesn't need humans anymore and wants to get rid of us?

These are all fair questions, and the concerns they raise need to be taken seriously. But there's a good reason to think that we can deal with them: This is not the first time a major innovation has introduced new threats that had to be controlled. We've done it before.

Whether it was the introduction of cars or the rise of personal computers and the Internet, people have managed through other transformative moments and, despite a lot of turbulence, come out better off in the end. Soon after the first automobiles were on the road, there was the first car crash. But we didn't ban cars—we adopted speed limits, safety standards, licensing requirements, drunk-driving laws, and other rules of the road.

We're now in the earliest stage of another profound change, the Age of AI. It's analogous to those uncertain times before speed limits and seat belts. AI is changing so quickly that it isn't clear exactly what will happen next. We're facing big questions raised by the way the current technology works, the ways people will use it for ill intent, and the ways AI will change us as a society and as individuals.

In a moment like this, it's natural to feel unsettled. But history shows that it's possible to solve the challenges created by new technologies.

I have written before about how AI is going to revolutionize our lives. It will help solve problems—in health, education, climate change, and more—that used to seem intractable. The Gates Foundation is making it a priority, and our CEO, Mark Suzman, recently shared how he's thinking about its role in reducing inequity.

I'll have more to say in the future about the benefits of AI, but in this post, I want to acknowledge the concerns I hear and read most often, many of which I share, and explain how I think about them.

One thing that's clear from everything that has been written so far about the risks of AI—and a lot has been written—is that no one has all the answers. Another thing that's clear to me is that the future of AI is not as grim as some people think or as rosy as others think. The risks are real, but I am optimistic that they can be managed. As I go through each concern, I'll return to a few themes:

- Many of the problems caused by AI have a historical precedent. For example, it will have a big impact on education, but so did handheld calculators a few decades ago and, more recently, allowing computers in the classroom. We can learn from what's worked in the past.
- Many of the problems caused by AI can also be managed with the help of AI.
- We'll need to adapt old laws and adopt new ones—just as existing laws against fraud had to be tailored to the online world.

In this post, I'm going to focus on the risks that are already present, or soon will be. I'm not dealing with what happens when we develop an AI that can learn any subject or task, as opposed to today's purpose-built AIs. Whether we reach that point in a decade or a century, society will need to reckon with profound questions. What if a super AI establishes its own goals? What if they conflict with humanity's? Should we even make a super AI at all?

But thinking about these longer-term risks should not come at the expense of the more immediate ones. I'll turn to them now.

Deepfakes and misinformation generated by AI could undermine elections and democracy.

The idea that technology can be used to spread lies and untruths is not new. People have been doing it with books and leaflets for centuries. It became much easier with the advent of word processors, laser printers, email, and social networks.

AI takes this problem of fake text and extends it, allowing virtually anyone to create fake audio and video, known as deepfakes. If you get a voice message that sounds like your child saying "I've been kidnapped, please send $1,000 to this bank account within the next 10 minutes, and don't call the police," it's going to have a horrific emotional impact far beyond the effect of an email that says the same thing.

On a bigger scale, AI-generated deepfakes could be used to try to tilt an election. Of course, it doesn't take sophisticated technology to sow doubt about the legitimate winner of an election, but AI will make it easier.

There are already phony videos that feature fabricated footage of well-known politicians. Imagine that on the morning of a major election, a video showing one of the candidates robbing a bank goes viral. It's fake, but it takes news outlets and the campaign several hours to prove it.

How many people will see it and change their votes at the last minute? It could tip the scales, especially in a close election.

When OpenAI co-founder Sam Altman testified before a U.S. Senate committee recently, Senators from both parties zeroed in on AI's impact on elections and democracy. I hope this subject continues to move up everyone's agenda.

We certainly have not solved the problem of misinformation and deepfakes. But two things make me guardedly optimistic. One is that people are capable of learning not to take everything at face value. For years, email users fell for scams where someone posing as a Nigeran prince promised a big payoff in return for sharing your credit card number. But eventually, most people learned to look twice at those emails. As the scams got more sophisticated, so did many of their targets. We'll need to build the same muscle for deepfakes.

The other thing that makes me hopeful is that AI can help identify deepfakes as well as create them. Intel, for example, has developed a deepfake detector, and the government agency DARPA is working on technology to identify whether video or audio has been manipulated.

This will be a cyclical process: Someone finds a way to detect fakery, someone else figures out how to counter it, someone else develops counter-countermeasures, and so on. It won't be a perfect success, but we won't be helpless either.

AI makes it easier to launch attacks on people and governments.

Today, when hackers want to find exploitable flaws in software, they do it by brute force—writing code that bangs away at potential weaknesses until they discover a way in. It involves going down a lot of blind alleys, which means it takes time and patience.

Security experts who want to counter hackers have to do the same thing. Every software patch you install on your phone or laptop represents many hours of searching, by people with good and bad intentions alike.

AI models will accelerate this process by helping hackers write more effective code. They'll also be able to use public information about individuals, like where they work and who their friends are, to develop phishing attacks that are more advanced than the ones we see today.

The good news is that AI can be used for good purposes as well as bad ones. Government and private-sector security teams need to have the latest tools for finding and fixing security flaws before criminals can take advantage of them. I hope the software security industry will expand the work they're already doing on this front—it ought to be a top concern for them.

This is also why we should not try to temporarily keep people from implementing new developments in AI, as some have proposed. Cyber-criminals won't stop making new tools. Nor

will people who want to use AI to design nuclear weapons and bioterror attacks. The effort to stop them needs to continue at the same pace.

There's a related risk at the global level: an arms race for AI that can be used to design and launch cyberattacks against other countries. Every government wants to have the most powerful technology so it can deter attacks from its adversaries. This incentive to not let anyone get ahead could spark a race to create increasingly dangerous cyber weapons. Everyone would be worse off.

That's a scary thought, but we have history to guide us. Although the world's nuclear nonproliferation regime has its faults, it has prevented the all-out nuclear war that my generation was so afraid of when we were growing up. Governments should consider creating a global body for AI similar to the International Atomic Energy Agency.

AI will take away people's jobs.

In the next few years, the main impact of AI on work will be to help people do their jobs more efficiently. That will be true whether they work in a factory or in an office handling sales calls and accounts payable. Eventually, AI will be good enough at expressing ideas that it will be able to write your emails and manage your inbox for you. You'll be able to write a request in plain English, or any other language, and generate a rich presentation on your work.

As I argued in my February post, it's good for society when productivity goes up. It gives people more time to do other things, at work and at home. And the demand for people who help others—teaching, caring for patients, and supporting the elderly, for example—will never go away. But it is true that some workers will need support and retraining as we make this transition into an AI-powered workplace. That's a role for governments and