Picture credit: US Dept Defense

PROS AND CONS

Neil Roseman *explains how AI innovation can come with great rewards and grave risks*

It can be difficult to see the risks that come along with the rewards of new technology. Digital innovations often seem to be blind to the potential technology and security risks that invariably come with early-generation technology, no matter how dazzling the potential of that technology. We see this over and over again as new technologies make their way into popular use. The IoT, for example, has transformative potential over countless sectors, but can only do so when given control of key operations within that sector. As such, the scope for failure is widened and the potential damage deepened. We see these risks wherever the IoT has taken meaningful root: industry, transportation, healthcare and city management.

We are now seeing the same with the emergence of Artificial Intelligence. It seems potentially applicable to every sector and has the potential to transform many aspects of our lives. Generative AIs like ChatGPT are already beginning to improve certain sectors in unexpected ways and as development continues that's likely to expand further.

The piloting of drones is already being handed over to AI models and automated systems

AI is already being given responsibility for sensitive data and crucial operations. Generative AI, for example, has quickly become an indispensable tool for software development, helping programmers meet ever increasing demand for new apps and services. However, it is are introducing novel risks.

AI LEARNS AND ADAPTS AS IT INTERACTS WITH THE WORLD AND DIFFERENT SYSTEMS IT CONNECTS TO

AI provides key insights into how to write code quickly, but doesn't necessarily write it better. There have been multiple recorded examples of a generative AI giving out mistaken or bad advice. At scale, this could lead to huge risks through technology failures, security risks and ultimately quickly amounting security debt from misinformed decision making in software development. One study from Stanford University recently found that AI tools actually produced worse results in coding. The paper's authors concluded: "We observed that participants who had access to the AI assistant were more likely to introduce security vulnerabilities for the majority of programming tasks, yet were also more likely to rate their insecure answers as secure compared with those in our control group."

AI can also be manipulated by prompt injection attacks. There is now a thriving community that researches and documents ways to manipulate an AI application by simply asking it a series of questions, which allows a malicious party to manipulate the application's output. This can result in the exposure of sensitive data, the breaching of the AI's internal rules or otherwise sewing mayhem in the target systems.

It's not just that it is offering bad data and mistaken insights, it's also effectively an attack vector themselves. AI applications – especially generative AI – often sit in front of sensitive data and systems. When they're breached, attackers can access those data and systems or otherwise manipulate the application which lies on top of them.

The development of AI is a complex and arduous process, full of complications which can spring up at any time. The troubling reality is that as AI interacts with other systems and data, it can react in unknown ways. Which brings us to some very important considerations when we think about the potentially transformative use cases in which AI will be used.

HEALTHCARE

The Healthcare sector is widely touted as one of the areas which stands to gain most from AI. Statista has revealed that the AI healthcare market will be worth $187-billion by 2030. It's hard not to see why — its applications are wide within the sector, covering everything from mere administration and to genomic analysis, biomedical research, medical imaging and clinical diagnosis. The Harvard School of Public Health predicts that using AI as a diagnostic tool could lower treatment costs by 50 percent while improving health outcomes by 40 percent. Indeed, the AI revolution in healthcare is already well underway.

Whatever the promise of this technology, it also opens up a new category of grave risks. Hospitals and healthcare organisations have been prime targets for cyber attacks over the last few years. One study from the University of Minnesota School of Public Health found that during ongoing ransomware attacks, patient mortality increased significantly. Furthermore, the study estimates that between 2016 and 2021, ransomware attacks on hospitals contributed to the deaths of between 42 and 67 people.

The responsibility of AI systems here is great. They will be proving the insights that help medical professionals make life changing decisions for patients. The overwhelming need for precision in those devices means the data that trains those models must be collected and handled with the utmost care.

Similarly, medical devices can become attack vectors. The now-famous hackable insulin pump is just such an example. An AI model in medical settings could be manipulated to even more catastrophic ends.

MILITARY DRONES

The battlefield revolution of the last generation is the drone. These are now found throughout nearly every battlefield on earth. The Unmanned Aerial Vehicle (UAV) as it is sometimes known, is often piloted by humans but looks like a perfect fit for AI technology. Yet, much like healthcare, in doing so it will be charged with a huge responsibility. Indeed, this is already happening and the piloting of drones is already being handed over to AI models and automated systems.

The people now using these tools are now AI and ML specialists. The arms industry has always been a major target for international espionage and it seems likely that AI systems will be a target too. There are a number of ways to corrupt these systems considering the sheer level of precision required of them. If for example, the training data for these systems were corrupted or a well-resourced attacker launched a data poisoning attack against these systems, the resulting accidents could have a massive effect on the company and country using them.

AUTONOMOUS VEHICLES

Perhaps the most vaunted case of the possibilities of AI is the autonomous vehicle. Again, the same risks are emerging. As with healthcare and military armaments, transportation is a safety-critical field. An autonomous vehicle will have to accommodate and adapt to the potential millions of conditions that it will face on the road. That means a mass of clear, performant data which can be deployed to teach an autonomous vehicle exactly how it will need to operate in the real world. Errors and biases in that data will have serious consequences for passenger safety.

There are also potential security risks that emerge. There have already been multiple examples of exploits against vehicles and the connected systems of an autonomous vehicle will need to be highly resilient, lest they become a vector to access and manipulate the AI models which are running the autonomous vehicle.

These are just a few examples of the ways in which AI may revolutionise a variety of sectors while simultaneously introducing risk into those use cases in which it is deployed.

To state the obvious: AI is a complex technology. The supply chain of an AI model is worthy of particular focus. AIs sit at the end of certain supply chains, the beginnings of others and right in the middle of others. This creates a

galaxy of points of failure for an AI model and the risk of a mistake in one place, scaling as that AI model is further employed. Take the example of software development in which AIs are used to help write applications and tools which will go on to help develop other AI models. We give these AI tools incredible trust and yet according to a number of security researchers, they commonly give poor advice, which leads to vulnerabilities and code errors. Taken together, it's not hard to see technical debt accumulating chaotically and widely in the coming years.

UNMANNED AERIAL VEHICLES LOOK LIKE A PERFECT FIT FOR ARTIFICIAL INTELLIGENCE

AI models are often cobbled together from a number of different pre-fabricated items, open source components and data. These can, and often do, introduce a large amount of risk.

A study from North Carolina State University found that the Deep Neural Networks (DNN) used within AI models were full of vulnerabilities that could allow the end product to be maliciously manipulated. One of the authors later stated that: "Attackers can take advantage of these vulnerabilities to force the AI to interpret the data to be whatever they want. This is incredibly important because if an AI system is not robust against these sorts of attacks, you don't want to put the system into practical use – particularly for applications that can affect human lives."

Similarly, late in 2023, a critical vulnerability was found in the open-source TorchServe machine learning frameworks that many AI models are built with. That critical vulnerability would allow an attacker to access the models proprietary data and insert their own malicious models into production. Huntr, the bug bounty platform that discovered the vulnerabilities, noted that it actually commonly finds critical vulnerabilities like these in the open source frameworks used for AI development.

Training data is also an area of particular sensitivity. As the data which initially teaches an AI model what to do, it can introduce grave problems. Data could be of poor quality, inaccurate or biased and thus harm the secure and reliable operation of the model with which it is built. It can also just be maliciously poisoned and it's important to point out that OWASP believes this to be one of the top current risks to LLMs and AIs.

One of the things that generates excitement and worry around AI is that so much of the technology around it is a black box. AI is a technology that learns and adapts as it interacts with the world and the systems it connects to. We ultimately don't know what the outcomes of many AI technologies will be. It's out of that black box that we can speculate endlessly about utopian successes and apocalyptic failures, but gain little concrete certainty about what will happen.

However, we can confidently talk about our inability to know these things. The fact remains that AI is a hugely complex technology, and that complexity only grows when we consider the individual use cases that it will be put into. It's also important to note that the people using and administering this technology will often not be specialists in the area, meaning that they may not know what to do in the case of a technology failure or cyber attack.

In short, this bold new frontier can be a dangerous place, full of pitfalls and threats. The international community of businesses, governments and the sectors that stand to gain the most from AI need to be wary of rushing headlong into new technological opportunities – exciting though they may be.

The benefit and applicability of one technology often runs in direct tension with the risk it provides. In this sense, it might be better – instead of risks and rewards – to talk about the responsibilities we have. The promises of AI might be exciting, but they'll never be fully captured without also taking serious responsibility for the risks ●

Neil Roseman is CEO of Invicti Security and Advisory Partner at Summit Partners, with over 20 years of experience in building high-scale software and web services for consumers and the enterprise.

Autonomous military vehicles have to accommodate and adapt to the potential millions of conditions that they will face on the road