ARTIFICIAL INTELLIGENCE, BIG DATA, SECURITY

The Cyber Security Battlefield

ROBERT FAY, WALLACE (WALLY) TRENHOLM

Artificial intelligence (AI) is truly a revolutionary feat of computer science, set to become a core component of all modern software over the coming years and decades. This presents a threat but also an opportunity. AI will be deployed to augment both defensive and offensive cyber operations. Additionally, new means of cyber attack will be invented to take advantage of the particular weaknesses of AI technology. Finally, the importance of data will be amplified by AI's appetite for large amounts of training data, redefining how we must think about data protection. Prudent governance at the global level will be essential to ensure that this era-defining technology will bring about broadly shared safety and prosperity.

■ ■ ■

AI and Big Data

In general terms, AI refers to computational tools that are able to substitute for human intelligence in the performance of certain tasks. This technology is currently advancing at a breakneck pace, much like the exponential growth experienced by database technology in the late twentieth century. Databases have grown to become the core infrastructure that drives enterprise-level software. Similarly, most of the new value added from software over the coming decades is expected to be driven, at least in part, by AI.

Within the last decade, databases have evolved significantly in order to handle the new phenomenon dubbed "big data." This refers to the unprecedented size and global scale of modern data sets, largely gathered from the computer systems that have come to mediate nearly every aspect of daily life. For instance, YouTube receives over 400 hours of video content each minute (Brouwer 2015).

> For instance, researchers have trained computer models to identify an individual's personality traits more accurately than their friends based exclusively on what Facebook posts they had liked.

Big data and AI have a special relationship. Recent breakthroughs in AI development stem mostly from "machine learning." Instead of dictating a static set of directions for an AI to follow, this technique trains AI by using large data sets. For example, AI chatbots can be trained on data sets containing text recordings of human conversation collected from messenger apps to learn how to understand what humans say, and to come up with appropriate responses (Pandey 2018). One could say that big data is the raw material that fuels AI algorithms and models.

The main constraint on innovation is no longer the difficulty in recording and storing information, but the finding of useful insights among the sheer abundance of data now being collected. AI can notice patterns in mammoth data sets that are beyond the ability of human perception to detect. In this way, the adoption of AI technology can make even mundane and seemingly trivial data valuable. For instance, researchers have trained computer models to identify an individual's personality traits more accurately than their friends can, based exclusively on what Facebook posts the individual had liked (Wu, Kosinski and Stillwell 2015).

AI and Cyber Security

Hardly a day passes without a news story about a high-profile data breach or a cyber attack costing millions of dollars in damages. Cyber losses are difficult to estimate, but the International Monetary Fund places them in the range of US$100–$250 billion annually for the global financial sector (Lagarde 2012). Furthermore, with the ever-growing pervasiveness of computers, mobile devices, servers and smart devices, the aggregate threat exposure grows each day. While the business and policy communities are still struggling to wrap their heads around the cyber realm's newfound importance, the application of AI to cyber security is heralding even greater changes.

One of the essential purposes of AI is to automate tasks that previously would have required human intelligence. Cutting down on the labour resources an organization must employ to complete a project, or the time an individual must devote to routine tasks, enables tremendous gains in efficiency. For instance, chatbots can be used to field customer service questions, and medical assistant AI can be used to diagnose diseases based on patients' symptoms.

In a simplified model of how AI could be applied to cyber defence, log lines of recorded activity from servers and network components can be labelled as "hostile" or "non-hostile," and an AI system can be trained using this data set to classify future observations into one of those two classes. The system can then act as an automated sentinel, singling out unusual observations from the vast background noise of normal activity.

Automating tasks that previously would have required human intelligence, such as using chatbots to field customer service questions, is one of the essential purposes of AI, and enables tremendous gains in efficiency for organizations. (Photo: Piotr Swat / Shutterstock.com)

This kind of automated cyber defence is necessary to deal with the overwhelming level of activity that must now be monitored. We have passed the level of complexity at which defence and identification of hostile actors can be performed without the use of AI. Going forward, only systems that apply AI to the task will be able to deal with the complexity and speed found in the cyber security environment.

Continuously retraining such AI models is essential, since just as AI is used to prevent attacks, hostile actors of all types are also using AI to recognize patterns and identify the weak points of their potential targets. The state of play is a battlefield where each side is continually probing the other and devising new defences or new forms of attack, and this battlefield is changing by the minute.

Perhaps the most effective weapon in a hacker's arsenal is "spear phishing" — using personal information gathered about an intended target to send them an individually tailored message. An email seemingly written by a friend, or a link related to the target's hobbies, has a high chance of avoiding suspicion. This method is currently quite labour intensive, requiring the would-be hacker to manually conduct detailed research on each of their intended targets. However, an AI similar to chatbots could be used to automatically construct personalized messages for large numbers of people using data obtained from their browsing history, emails and tweets (Brundage et al. 2018, 18). In this way, a hostile actor could use AI to dramatically scale up their offensive operations.

AI can also be used to automate the search for security flaws in software, such as "zero-day vulnerabilities." This can be done with either lawful or criminal intent. Software designers could use AI to test for holes in their product's security, just as criminals search for undiscovered exploits in operating systems.

AI will not only augment existing strategies for offence and defence, but also open new fronts in the battle for cyber security as malicious actors seek ways to exploit the technology's particular weaknesses (ibid., 17). One novel avenue of attack that hostile actors may use is "data poisoning." Since AI uses data to learn, hostile actors could tamper with the data set used to train the AI in order to make it do as they please. "Adversarial examples" could provide another new form of attack. Analogous to optical illusions, adversarial examples consist of modifying an AI's input data in a way that would likely be undetectable to a human, but is calculated to cause the AI to misclassify the input in a certain way. In one widely speculated scenario, a stop sign could be subtly altered to make the AI system controlling an autonomous car misidentify it as a yield sign, with potentially deadly results (Geng and Veerapaneni 2018).

The New Value of Data

AI technology will alter the cyber security environment in yet another way as its hunger for data changes what kind of information constitutes a useful asset, transforming troves of information that would not previously have been of interest into tempting targets for hostile actors.

While some cyber attacks aim solely to disrupt, inflict damage or wreak havoc, many intend to capture strategic assets such as intellectual property. Increasingly, aggressors in cyberspace are playing a long-term game, looking to acquire data for purposes yet unknown. The ability of AI systems to make use of even innocuous data is giving rise to the tactic of "data hoovering" — harvesting whatever information one can and storing it for future strategic use, even if that use is not well defined at present.

A recent report from *The New York Times* illustrates an example of this strategy in action (Sanger et al. 2018). The report notes that the Chinese government has been implicated in the theft of personal data from more than 500 million customers of the Marriott hotel chain. Although commonly the chief concern regarding data breaches is the potential misuse of financial information, in this case the information could be used to track down suspected spies by examining travel habits, or to track and detain individuals to use them as bargaining chips in other matters.

Data and AI connect, unify and unlock both intangible and tangible assets; they shouldn't be thought of as distinct. Quantity of data is becoming a key factor to success in business, national security and even, as the Cambridge Analytica scandal shows, politics. The Marriott incident shows that relatively ordinary information can now provide a strategic asset in the fields of intelligence and national defence, as AI can wring useful insights out of seemingly disparate sources of information. Therefore, this sort of bulk data will likely become a more common target for actors operating in this domain.

According to a report, the Chinese government has been implicated in the theft of personal data from over 500 million customers of the Marriott hotel chain using the tactic of data hoovering. (Photo: TK Kurikawa / Shutterstock.com)

Implications for Policy and Governance

These unfolding developments will force a rethinking of prevailing cyber security strategies. In an increasingly interconnected system, identifying the weakest link becomes more challenging, but also more essential. As sensors, machines and people become interwoven providers of data for valuable AI systems, there will be a proliferation of entry points for cyber attacks. Cyber security requires a comprehensive strategy to minimize weakest links; a piecemeal approach to cyber policy will not work. Since the training data that feeds the most important and revolutionary AI technologies is global in scope, gathered from across many different countries, it is clear that governance at the national level alone will not suffice.

Global policy makers have begun turning their attention to the ramifications of widespread AI technology, and its effect on cyber security in particular. The Group of Seven (G7) turned its attention to the governance of AI during the 2018 summit in Charlevoix, Quebec, pledging to “promote human-centric AI” through appropriate investments in cyber security, while paying heed to privacy and personal information protection regarding the data that serves as the raw input for machine learning (G7 2018).

The application of AI technology to pre-existing cyber attack strategies such as spear phishing will both augment their effectiveness and — by circumventing labour constraints — expand the number of actors capable of undertaking them. This lends a greater urgency to existing efforts to create effective global governance in cyberspace and international data protection, such as the United Nations Group of Government Experts’ attempt to establish accepted norms of conduct.

> While commonly thought of as a threat to privacy, AI also has the potential to help preserve privacy and exert control over proprietary data and its derived assets.

The very same pieces of technology that enable more threatening types of cyber attack are also driving growth in the civilian economy and enabling more effective cyber defence. While commonly thought of as a threat to privacy, AI also has the potential to help preserve privacy and exert control over proprietary data and its derived assets. Policy makers will have to carefully consider how to regulate the use of these technologies, balancing the need to keep powerful weapons out of the hands of malicious actors without stifling innovation. It will be essential to harmonize such policies across national jurisdictions. Since hostile actors are capable of reaching across borders with stupendous ease, any country that unilaterally restricts the use and development of these technologies within its borders would be putting itself at a competitive disadvantage.

Moreover, as AI technology becomes more integrated into the general economy and civilian sphere, existing legal and normative frameworks may need to be adjusted to cover novel forms of attack such as data poisoning and adversarial examples. Up to this point, data theft has been the main concern in cyberspace. Going forward, hostile actors will likely try to gain access to databases not only to obtain their information, but also to alter and manipulate them. The legal definition of what constitutes a cyber attack may need to be amended to cover these novel threats (Brundage et al. 2018, 57).

AI algorithms learn from data to produce a valuable new prediction tool, and the output of AI can be separated from the original training data. Therefore, to truly control the data and its value, any assets that are produced from data must also be controlled. The infrastructure that allows the recording, storage and analysis of big data should be treated as an asset just like it is in any other sector. Furthermore, some sectors, such as finance, have systemic implications, and are even more important to protect due to third-party linkages. Governing institutions will need to continue to improve their security posture in these and many other areas, including identity fraud. Since the AI software used for attack purposes is capable of rapidly evolving, this is an ongoing requirement rather than a one-off investment.

Works Cited

Brouwer, Bree. 2015. "YouTube Now Gets Over 400 Hours of Content Uploaded Every Minute." Tubefilter, July 26. www.tubefilter.com/2015/07/26/youtube-400-hours-content-every-minute/.

Brundage, Miles, Shahar Avin, Jack Clark, Helen Toner, Peter Eckersley, Ben Garfinkel, Allan Dafoe, Paul Scharre, Thomas Zeitzoff, Bobby Filar, Hyrum Anderson, Heather Roff, Gregory C. Allen, Jacob Steinhardt, Carrick Flynn, Seán Ó hÉigeartaigh, Simon Beard, Haydn Belfield, Sebastian Farquhar, Clare Lyle, Rebecca Crootof, Owain Evans, Michael Page, Joanna Bryson, Roman Yampolskiy and Dario Amodei. 2018. *The Malicious Use of Artificial Intelligence: Forecasting, Prevention and Mitigation*. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf.

G7. 2018. "Charlevoix: Common Vision for the Future of Artificial Intelligence." https://g7.gc.ca/wp-content/uploads/2018/06/FutureArtificialIntelligence.pdf.

Geng, Daniel and Rishi Veerapaneni. 2018. "Tricking Neural Networks: Create Your Own Adversarial Examples." *Machine Learning @ Berkley* (blog), January 10. https://ml.berkeley.edu/blog/2018/01/10/adversarial-examples/.

Lagarde, Christine. 2012. "Estimating Cyber Risk for the Financial Sector." *IMFBlog*, June 22. https://blogs.imf.org/2018/06/22/estimating-cyber-risk-for-the-financial-sector/.

Pandey, Parul. 2018. "Building a Simple Chatbot from Scratch in Python (Using NLTK)." *Medium,* September 17. https://medium.com/analytics-vidhya/building-a-simple-chatbot-in-python-using-nltk-7c8c8215ac6e.

Sanger, David, Nicole Pelroth, Glenn Thrush and Alan Rappeport. 2018. "Marriott Data Breach Is Traced to Chinese Hackers as U.S. Readies Crackdown on Beijing." *The New York Times,* December 11. www.nytimes.com/2018/12/11/us/politics/trump-china-trade.html.

Wu, Youyou, Michal Kosinski and David Stillwell. 2015. "Computer-based personality judgments are more accurate than those made by humans." *Proceedings of the National Academy of Sciences* 112 (4): 1036–40. www.pnas.org/content/112/4/1036.

The opinions expressed in this article/multimedia are those of the author(s) and do not necessarily reflect the views of CIGI or its Board of Directors.

ABOUT THE AUTHORS

Robert Fay

Robert (Bob) Fay is the managing director of digital economy at CIGI. The research under his direction assesses and provides policy recommendations for the complex global governance issues arising from digital technologies.

Wallace (Wally) Trenholm

Wally Trenholm is a senior fellow with CIGI, where he contributes his expertise on artificial intelligence (AI), data governance and international security. He is CEO of Sightline Innovation and a software architect with over 25 years of experience in areas of operations including AI, data governance, distributed computing systems and sensing technologies, focusing on applications in commercial, legal, financial and government use cases.

Governing Cyberspace during a Crisis in Trust

In the Series

AFRICA (91)
ARTIFICIAL INTELLIGENCE (95)
BIG DATA (205)
CENTRAL BANKING (105)
CHINA (223)
DEMOCRACY (233)
DIGITAL CURRENCY (32)
EMERGING TECHNOLOGY (123)
FINANCIAL SYSTEMS (316)
FUTURE OF WORK (33)
G20/G7 (301)
GENDER (91)
IMF (214)
INDIA (55)
INNOVATION (188)
INNOVATION ECONOMY (179)
INTELLECTUAL PROPERTY (169)
INTERNET GOVERNANCE (201)
INVESTOR STATE ARBITRATION (62)
MONETARY POLICY (123)
NAFTA/CUSMA (200)
PATENTS (20)
PLATFORM GOVERNANCE (330)
PRODUCTIVITY (12)
SECURITY (223)
SOVEREIGN DEBT (100)
STANDARDS (30)
SURVEILLANCE & PRIVACY (141)
SYSTEMIC RISK (24)
TRADE (499)
WTO (137)

Get regular updates on our research and events in your inbox.

Your email ✉ SIGN UP

Contact Careers Directions Privacy Notice Media