WILL ARTIFICIAL INTELLIGENCE HONE NORTH KOREA'S CYBER "ALL-PURPOSE SWORD"?

Scott W. Harold, Nathan Beauchamp-Mustafaga, Jenny Jun, Diana Myers, and Derek Grossman

How is the increasing spread of artificial intelligence (AI) likely to shape the cyber capabilities of the Democratic People's Republic of Korea (DPRK; North Korea) in the coming years?[1] Over the past decade, cyber tools have become an important enabler of the Kim Jong Un regime's quest to achieve its policy objectives.[2] Today, as a result of sustained investments by Pyongyang, the DPRK has developed an increasingly sophisticated set of cyber capabilities, which it has used to substantial effect against foreign militaries, banks, companies, media outlets, and individuals. While the regime has been able to achieve much through cyberattacks relying on traditional human operators, there are some areas where automating cyberattacks may prove attractive to the North.[3] At the same time, demand for trained cybersecurity professionals usually far outstrips supply, and AI for cyber defense may be an area where the regime ultimately feels compelled to invest, either to offset human capital shortfalls, or as adversary AI-enabled cyberattacks grow more sophisticated.[4] Is North Korea about to make the leap to AI-enabled cyberattacks or cyber defense?[5] Or will the country's international isolation impair its ability to pair up AI with its existing cyber capabilities?

Recognizing the challenges posed by Pyongyang's computer network operations, official national security documents from the United States, the Republic of Korea (ROK; South Korea), and Japan have all identified the DPRK's cyber operations as a serious concern. For example, the *2021 Worldwide Threat Assessment* by the U.S. Director of National Intelligence noted that North Korea's cyber program "poses a growing espionage, theft, and attack threat."[6] For its part, the ROK Ministry of National Defense (MND), in its *2020 Defense White Paper,* noted that North Korea is "operating a 6,800-strong unit of trained cyber-warfare specialists and is working to enhance cyber capabilities by continuing R&D on [the] latest technologies."[7] In 2020, Japan's Ministry of Defense concluded that North Korea possesses "large-scale cyber units as part of its asymmetric military capabilities, engaging in theft of military secrets and developing capabilities to attack critical infrastructure of foreign countries."[8] And as the ROK's MND has further noted, the North is also "fostering specialists and continuing R&D in [the] latest technologies" related to cyber technologies.[9] One such "latest technology" is AI: the use of computer software programs to sift through large volumes of data to identify patterns, predict behavior or results, and adjust and improve its predictions of outcomes in terms of data feedback.

Scott W. Harold is a Senior Political Scientist, ***Nathan Beauchamp-Mustafaga*** is an Associate Policy Researcher, and ***Derek Grossman*** is a Senior Defense Analyst at The RAND Corporation. ***Jenny Jun*** is a Nonresident Fellow at the Atlantic Council's Cyber Statecraft Initiative and Ph.D. candidate at Columbia University's Department of Political Science. ***Diana Myers*** is an officer in the United States Air Force and a Ph.D.-candidate at the Pardee RAND Graduate School. To read other KEI Special Reports, please visit https://keia.org/keia_publication/special-reports/

There are a number of possibilities for how AI might relate to North Korea's cyber capabilities. For example, if Pyongyang augments its cyberattacks with AI, the North might be able to rapidly accelerate and expand its intrusion sets by using algorithms to identify weaknesses in adversary systems or improve the effectiveness of its attacks. On the other hand, U.S., South Korean, or other nations that employ AI for cyber defense may become more proficient at detecting and defeating North Korea's human-developed cyber intrusion sets, eroding the value of Pyongyang's cyber arsenal unless it improves its offensive cyber tactics, techniques, and procedures (TTP), possibly by employing AI for offensive cyber in novel ways. The regime could also seek to employ AI to improve its own cyber defenses, hoping to detect and defeat the United States', South Koreans', or other nations' efforts to probe or penetrate the limited systems that actors in the North use to connect to the Internet. Finally, and in response to the automation of its own cyberattacks or cyber defenses, Pyongyang might target adversary AI training data or models themselves.

To date, we find that there is very limited direct evidence that the DPRK has moved to pair AI with its cyber capabilities, but compelling logic and significant circumstantial evidence indicate that it will do so in the years ahead. Given the nature of the DPRK as a closed "hard target" country, with the details of its cyber programs as presumably among the most closely-held secrets of state, a lack of direct evidence about its cutting-edge capabilities is hardly diagnostic. As has been shown time and again by progress in North Korea's strategic weapons programs, Pyongyang's technical knowledge base or intent to develop an advanced capability can be good markers of its ultimate goals. We did find significant circumstantial evidence, precedent, and logic that points to the possibility that North Korea has, or in the future likely will, pair its cyber capabilities with AI. Meanwhile, we found no evidence or compelling logic to support hypotheses that the DPRK would choose not to pursue such capabilities, or that the DPRK would be unable to develop and employ such capabilities.

In the absence of concrete evidence, there are very good reasons to think that the regime is moving in this direction based on its interests; trends in broader technology, espionage, and warfare; statements by regime leaders; a survey of academic writings on AI by DPRK researchers; as well as past evidence from how the North has embraced other strategic technologies and how other nations are treating AI and cyber. While Pyongyang's human-conducted cyberattacks have been quite effective to date, there are nonetheless some good reasons to think that in the future it may seek to incorporate AI its offensive cyber operations, especially if its targets begin to degrade the effectiveness of Pyongyang's cyberattacks by adopting AI-enabled anomaly detection.[10] And though the North is connected to the global Internet through only a fairly limited number of access points today, should its own economy develop greater touchpoints with the outside internet, the North would likely see increased value in AI-enabled cyber defenses. Finally, an examination of the experiences of North Korea and other similar malign actors finds no evidence that the cost of AI development, access to cutting edge research or training sets, or availability of electricity supply are likely to constrain the regime should it choose to develop AI for cyber, while access to talent and computing power could be somewhat more substantial chokepoints in AI development, though the former would potentially just further incentivize the North to press ahead even faster.

The remainder of this article unfolds in three parts.

First, we ground our discussion in an explanation of North Korea's overall policy goals and its political-military strategy. We then describe what is known or believed to be true about the role of cyber tools in supporting the regime's goals.

Second, we lay out what we know or can reasonably infer about North Korea's interest in and access to AI and machine learning (ML), as well as how these appear to fit with its overall cyber strategy. To characterize the DPRK's ambitions and capabilities in this area, we look at North Korean leadership statements and North Korean efforts to develop other strategic technologies. We then supplement these with an examination of insights drawn from a novel dataset we built of forty-eight technical articles written by researchers at Kim Il Sung University, the premiere school for the study of AI in North Korea, published between 2018 and 2020. Additionally, we compare North Korea's situation with that of other countries seeking controlled strategic technologies and assess how a series of factors may or may not constrain the regime's adoption of AI-enabled cyber.

Finally, the article explores how North Korea's cyber capabilities might evolve if combined with AI over the coming half-decade.

PYONGYANG'S POLICY GOALS, POLITICAL-MILITARY STRATEGY, AND THE ROLE OF CYBER

Since the founding of the DPRK in 1948, North Korea has pursued a highly consistent set of goals across the three regimes of Kim Il Sung, Kim Jong Il, and Kim Jong Un. Scholars assess that the DPRK's three fundamental goals are: 1) preserving the Kim family regime's control over the Northern half of the peninsula; 2) weakening and ultimately breaking the U.S.-ROK alliance; and 3) achieving the conquest of the rival South Korean regime, by force if necessary.[11] Official U.S. Department of Defense assessments similarly find that the regime's "overriding strategic goal" is to ensure the Kim family's perpetual rule of North Korea, followed by "reunification with the ROK, by force if necessary, [which] is a key component of North Korea's national identity, validating its policies and strategies, and justifying the sacrifices demanded of the populace."[12]

Since the 1990s, the DPRK's political-military strategy has tended to focus on efforts to magnify the regime's ability to asymmetrically hurt or coerce its neighbors and adversaries, implying that it can be induced to return to a more passive and less threatening posture only if bought off with concessions. At times, observers have confused this risk manipulation-centric political-military strategy with aspects of the regime's routine weapons development and testing programs designed to modernize and improve its military capabilities. The two, however, are logically distinct, even if the regime frequently seeks to extract value in the form of side payments from actors alarmed when it takes actions designed to test or advance its military capabilities.[13] Recognizing that the regime can ill-afford to trade space for time in an actual conflict, Pyongyang has typically pursued only local, limited military operations designed to advance a specific and constrained political goal, often paired with or augmented by the regime's sophisticated political warfare operations.[14] Owing to the difficulty of detecting and countering cyber intrusions, as well as attributing them reliably to the North, the regime sees cyber operations as particularly appealing tools to leverage in such political-diplomatic campaigns.

In order to understand the evolution of cyber within North Korea's overall national security toolkit, it is necessary to bear in mind that North Korea has long relied on asymmetric and unconventional means to make strategic gains against its adversaries. In this context, cyber capabilities have provided North Korea several attractive features compared to preexisting means. As Jung Pak, a former Deputy National Intelligence Officer for North Korea and currently Deputy Assistant Secretary of State, has noted, the senior leadership of the DPRK have directed the use of the regime's cyber tools to "coerce, conduct espionage, and earn currency for the regime."[15] Additional, Pak points out, in 2003 Kim Jong Il reportedly advised his senior military leadership that information had replaced bullets and oil as the key strategic driver of 21st Century warfare.

For his part, Kim Jong Un is reported to have referred to cyber as an "all-purpose sword," one that can be used for intelligence-gathering, revenue generation, cyber-enabled economic warfare, or in support of military operations.[16] No surprise then that Anna Fifield of *The Washington Post* reports that North Korean university students who have been through the regime's cyber training program describe hacking as "the country's strongest weapon," one reportedly used against South Korea as much as 1.5 million times daily.[17]

Offensive cyber capabilities allow North Korea to project power in a cost-effective manner even under geographic isolation, especially against states that traditionally maintained strategic depth. While its effects are not wholly substitutable, cyber capabilities can achieve some destructive and disruptive effects at lower cost, compared to the time and resources allocated to developing and maintaining a missile program or a large special operations force. For North Korea, offensive cyber capabilities help the regime overcome its geographic constraints on power projection and give it the power to harm targets in the continental United States directly, and therefore enhance its bargaining position in the Korean peninsula. Indeed, the regime has continued and accelerated its use of cyberattacks against the South, even during periods when it was negotiating directly with the progressive Moon Jae-in administration in 2018 and since.[18]

By contrast, defensive cyber capabilities are probably somewhat less of a concern for Pyongyang; one interviewee we spoke with characterized the North's connectedness to the global internet as a "pinhole."[19] While interconnectedness and digital information have become important underpinnings of daily activity for most countries around the world, the North Korean citizenry and the portions of the North Korean economy not directly tied to regime financing via criminal cyber activity are still far less connected to and reliant on cyberspace than is true of any of the DPRK's adversaries. With the U.S. and its allies already faced with limited levers to influence North Korea's behavior, this asymmetric dependence on cyberspace has created new opportunities for North Korea to seek gains with relatively little concern for retaliation in kind.

Additionally, in the cybercrime domain, North Korea's extensive cyber operations provide vital funds for propping up the Kim family rule and undermining the international community's sanctions regime, one of the few levers for influencing North Korea's behavior. The ability to simply steal and extort money from victims around the world not only feeds back into North Korea's weapons programs, but also helps to concentrate wealth among the elites rather than in the hands of the citizens. One expert we spoke with called this "the major shift from 2015 to the present" in the regime's use of cyber, namely "from strategically motivated hacking to pure profit generation and bank heists."[20] Indictments unsealed by the U.S. Department of Justice in early 2021 reinforce just how central cyber has become to the regime's finances in recent years, listing cyber-enabled bank heists, ATM cash-out thefts, extortion and ransomware, attacks on cryptocurrency exchanges, and development and deployment of malicious cryptocurrency applications, among other criminal actions leveraging the cyber domain to generate income.[21] Such an approach to revenue generation by the regime relieves pressures for domestic economic reforms that could potentially destabilize the current political structure. Ironically, such an approach could ultimately prove a vulnerability, since denial of access to cyberspace could prove a source of leverage against the regime if it becomes increasingly dependent on cybercrime resources.

Reliably describing the strength of a closed regime's cyber capabilities is challenging.[22] Cyber operations are often highly dependent on context and less consistently dependent on easily quantifiable metrics such as the number of "hackers" or "weapons" (a challenge not exclusive to the cyber domain). While task-oriented metrics such as average "breakout time"—the amount of time it takes for an adversary to move laterally inside a victim's network once it gains initial access—can be a more useful measure, even such measures are imperfect, since different Advanced Persistent Threat (APT) groups tend to target different categories of victims. For example, a 2019 industry report noted that North Korea came in second in average breakout time at 2 hours and 20 minutes behind Russia (19 minutes) and ahead of China (4 hours) and Iran (5 hours and 9 minutes), but it is not clear that these reported statistics were controlled for the victims' defense specifications, sector type, and purpose of intrusion.[23] In other words, rather than a reflection of an attacker's skills and capabilities, such statistics may simply reflect a choice of targeting focus and the average defense level of the systems the attacker is trying to penetrate. Furthermore, operational speed is only one metric, and is not necessarily more important than other possible goals an attacker might possess, such as stealth or concern for avoiding attribution or collateral damage.

Authoritative U.S. and South Korean government documents and statements by public officials have suggested that North Korea has roughly 6,000-7,000 cyber operators.[24] The Republic of Korea's 2020 Ministry of National Defense White Paper, for example, assessed that the North was actively training cyber warfare specialists and had developed a cadre of approximately 6,800 hackers.[25] It is worth noting that not all of these cyber specialists will likely be at the same skill level, with some elite operators and others likely less competent, a reality that further complicates any attempt to easily characterize the North's cyber operators.

In addition to these cyber operators, according to the UN Panel of Experts monitoring the enforcement of UN sanctions on the DPRK, it also has over 1,000 information technology workers employed outside the country in defiance of UN Security Council resolutions, earning foreign currency for the regime.[26] These workers also can provide the DPRK with know-how on emerging technologies like AI. The illicit presence of these workers overseas may also make it difficult to differentiate North Koreans who are merely illicitly earning money through otherwise legal means as opposed to cyber operators.[27]

According to a 2020 report, the U.S. Army estimates approximately 6,000 of the North's cyber specialists work for Bureau 121 of the Reconnaissance General Bureau (RGB), also known as the Cyber Warfare Guidance Unit. The RGB is the primary North Korean agency responsible for the regime's cyber activities, and was established around 2009.[28] It operates independently from the Korean People's Army (KPA) and reports directly to the State Affairs Commission, the highest decision-making body in North Korea, headed by Kim Jong Un. Many of Bureau 121's staff are believed to operate outside of North Korea.[29] In addition to the RGB, the General Staff Department of the Korean People's Army is also responsible for various cyber operations.

North Korean hackers in these agencies are recruited and trained from an early age through the regime's education system. According to a DPRK defector, talented youths are identified in early primary school and sent to special training schools at top North Korean universities such as Kim Chaek University of Technology, Mirim University, and the Kim Il Sung University.[30] Agencies such as the RGB then recruit the top graduates from these programs annually, with many sent abroad to China, Russia, or other countries for additional cyber training, and possibly even to operate as hired cybersecurity personnel for foreign firms.[31]

In terms of actual field operations, since 2009, North Korea has demonstrated that its cyber operations can be persistent, adaptive, and destructive. Table 1 lists a small sample of major cyberattacks that have been associated with the North.

Table 1. Selected Cyber Operations Publicly Attributed to North Korea[32]

Year	Name of Attack	Threat[33]	Tactics	Target	Country Hit
2009	4th of July Campaign	Denial of Service	Traffic Generation	Websites for ROK and U.S. presidential offices, defense, and other high-level institutions	ROK & U.S.
2011	Ten Days of Rain[34]	Denial of Service	Malware	Servers at ROK's highest government institutions and financial institutions	ROK
2013	DarkSeoul[35]	Denial of Service, Tampering	Malware, Remote Execution	ROK broadcasting stations and financial institutions	ROK
2013	Kimsuky[36] (first attribution)	Spoofing, Elevation of Privilege, Information Disclosure	Social Engineering/ Spear-phishing, Malware, Remote Execution, Data Exfiltration	ROK think tanks	ROK

Year	Name of Attack	Threat[33]	Tactics	Target	Country Hit
2014	Korean Hydro and Nuclear Power[37]	Information Disclosure	Social Engineering/ Phishing, Malware, Data Exfiltration	Korean Hydro and Nuclear Power (KHNP)	ROK
2014	Sony Hack[38]	Information Disclosure	Social Engineering/ Spear- phishing, Malware, Data Exfiltration	Sony Pictures Corporation	U.S.
2016	Military Plans[39]	Information Disclosure	Mobile Device Exploitation, Data Exfiltration	ROK Officials	ROK
2016	Bangladesh Bank Heist (FASTCash)[40]	Spoofing, Elevation of Privilege, Tampering	Spear-phishing, Watering Holes, Code Injection, Credential Harvesting, etc.	Bangladesh Bank officials' credentials	Bangladesh
2017	Youbit Hack[41]	Information Disclosure	Unknown (Credential Harvesting)	Youbit (ROK Bitcoin Exchange)	ROK
2017	WannaCry[42]	Denial of Service	Ransomware	Various targets' data was held for ransom; particularly severe damage to UK's National Health Service	Various
2018	Coinrail and Bithumb hacks[43]	Information Disclosure	Unknown (Credential Harvesting)	Coinrail lost $37 million while Bithumb lost $40 million in cryptocurrency	ROK
2018	GhostSecret[44]	Information Disclosure	Malware, Remote Execution, Data Exfiltration	Various targets; operation compromised firms in telecommunications, health, finance, entertainment, critical infrastructure	Estimated 17 countries

These attacks reveal three strategic goals of North Korea's cyberattacks: (1) to generate revenue, (2) to cause disruption or impose consequences on those who run afoul of the regime, and (3) to gather intelligence and conduct espionage.[45]

North Korea's cyber operations have also exhibited a blatant disregard for restraint, even going so far as to target South Korean civilian nuclear power plants.[46] North Korea's cybercrime activities have struck banks, cryptocurrency exchanges, and even individual ATMs, while its methods have ranged from outright theft to ransomware, blackmail using exfiltrated data or attempting to sell it on the black market, and even, in one instance, attempting to set up a cyber protection racket.[47] North Korea has even inserted destructive payloads in order to mask criminal activity and obfuscate investigation, for instance by using a wiper against a Chilean bank in 2018 to cover fraudulent SWIFT transactions. Examples such as WannaCry further indicate that North Korea has little regard for who it hurts with its indiscriminate targeting, though it is unclear whether this was evidence of incompetence, carelessness or deliberate malice.[49] Such brazen tactics set North Korean hackers apart from those of most other state-sponsored groups. Additionally, just as in the real world, so too in the virtual world the regime has ties to non-state criminal groups acting online, and has used these in a variety of ways, including to purchase access to victims who have already been penetrated by other cybercrime actors; by buying and selling stolen victim credentials on the illegal market; and by selling its services to others as hackers-for-hire.[50]

North Korean hackers are often successful at achieving the task at hand, often by using any means necessary. At the same time, they have also exhibited a lack of proficiency in specific areas such as cryptography, indicating that North Korea's cyber prowess should not be overestimated (or, alternatively, that it has room to improve further). For example, Kaspersky reported in July 2020 that North Korea was spreading a novel ransomware variant called VHD, but noted that several elements in its encryption process allowed for recovery of original data.[51] Similarly, a 2018 analysis of three North Korean indigenous encryption algorithms used for its Red Star operating system noted that they were simple variants of existing encryption schemes such as Advanced Encryption Standard and Secure Hash Algorithm-1 and had several operations that undermined its security, and kept it vulnerable to a side-channel attack.[52] Although North Korea claimed in 2017 to have advanced research in quantum cryptography for secure communications and two years later to have fielded its own indigenous encryption algorithm for Koryolink, the country's wireless telecommunications service provider, more evidence is needed regarding such claims in order to assess their reliability.[53]

As the foregoing review of North Korean cyber capabilities suggests, the network domain is one where the DPRK sees substantial advantages owing to reach, anonymity, and asymmetry. At the same time, trends in foreign technologies for cyber offense and cyberdefense—particularly those associated with automation and the merging of artificial intelligence and machine learning with computer network

operations—could carry substantial implications for North Korea's ability to continue employing its current approach to cyber. The next section examines the state of North Korean research on AI, recognizing that as foreign nations move to enhance the automation of their network operations, North Korea is likely to follow that trend and seek to do so as well.

DPRK AI CAPABILITIES

Discerning what North Korea knows, needs, has, and aspires to do with AI and machine learning as applied to cyber is a challenging question to answer with great confidence due to the aforementioned limitations on the availability of detailed information about the inner workings of North Korea's cyber program. This section, therefore, relies heavily on a foundation of the sorts of information the regime has allowed in public, and then applies logic and comparative cases to compensate for the information gaps. It starts by surveying North Korean leadership statements relating to advanced technology and AI/ML. We also identify and evaluate the state of North Korean education focused on AI/ML, focusing in particular on talent training and the technical writings of North Korean scholars on AI published in the *Kim Il Sung University Gazette,* one of the country's leading institutions for the study of AI. We also identify and assess the possible constraints on North Korea's development of AI, and close with a comparison of what the experiences of other countries seeking to acquire AI can tell us about North Korea's efforts to develop advanced AI.

North Korean Leadership Statements on Advanced Technologies and AI/ML

Despite its very limited information exchanges and points of contact between its own intranet and the outside world, North Korea has managed to develop a very useful—if not necessarily technologically cutting-edge—set of cyber capabilities. While its attributed intrusion sets have largely been limited to technically relatively unsophisticated attacks, its use of these has nonetheless been fairly effective, innovative, and adaptive.

Worryingly, evidence appears to suggest that the North may now be applying the same efforts to explore at least lower-level AI development as applied to cyber. Public media sources suggest that North Korea is centering the development of AI capabilities at the forefront of what the regime describes as its Fourth Industrial Revolution, a term the DPRK uses to describe Kim Jong Un's push to modernize the economy by focusing on mass communications, the Internet, and artificial intelligence.[54] A November 2, 2018 article in *Rodong Sinmun* stated that the Intelligent Technology Institute at Kim Il Sung University was "burning with ambition to hold supremacy in the artificial intelligence field…and to contribute to the establishment of the AI technology industry in the country."[55] The following year, a *Rodong Sinmun* article stated that "in the age of Artificial Intelligence, data is more valuable than gold or gas, therefore, collection and analysis of data is more critical than ever."[56] Perhaps reflecting an increased emphasis on AI, another article reported that Pyongyang University of Computer Science planned to change its computer-programming department into an AI department, though this could also be an exaggeration or a reflection of the pressures on institutions inside North Korea to demonstrate that they are supportive of the perceived preferred policy direction set by the country's leader.[57]

More recently, on August 4, 2021, *Rodong Sinmun* reported that "it is… the Party's intention to increase investment in science and technology and science research and guarantee the research and living conditions of scientists."[58] It added, "we must provide scientists and technicians all data needed for science research, including the latest scientific research, and actively strive to provide them the conditions and time to be able to master them."[59] Thus, despite the regime's political and economic isolation, North Korean AI researchers are being encouraged by Pyongyang, and official media reports signal that the North Korean leadership has placed an increasing emphasis on AI. This aligns with the Kim regime's support for scientists supporting other DPRK priorities, such as the country's missile program.[60]

North Korean Education and Research on AI/ML Technologies

North Korea's efforts to advance its capabilities in AI go beyond just leaders' statements and rhetoric; the regime is actively pursuing initiatives to advance its capabilities both at home and abroad through university training, foreign exchange programs, and other means. This section provides an overview of North Korea's state of academic research on AI/ML in order to assess North Korea's level of interest and proficiency in the topic.

Currently, there is no direct evidence in open sources that North Korea has already applied AI or ML to enhance its offensive cyber programs.[61] North Korea also has not published any military strategy or doctrine concerning cyber warfare, let alone any specific statements on applying artificial intelligence to it. It is possible, however, to gain insight into the North's AI capabilities and ambitions from UN reports as well as publications by leading North Korean universities on areas of AI-related academic research.

The *Panel of Experts* established pursuant to UN Security Council Resolution 1718 has been identifying North Korean sanctions violations and related suspicious activity in regular public reports since 2010, and these reports include information not only on North Korean research and IT activities, but also specifically examples related to AI. Most recently, the 2021 edition of the *Final Report* noted that professors at North Korea's Kim Chaek University had developed a technique to help characterize the strength of rock masses around a coal mine by analyzing images using an artificial neural network, a specific type of artificial intelligence.[62]

The most compelling evidence of North Korean AI development, however, comes from one of its own journals. Since 2018, the *Kim Il Sung University Gazette* has been publishing a new quarterly Information Sciences series as a separate discipline from the previous series on Natural Sciences. Between 2018 and 2020, a total of 48 articles were published pertaining to either machine learning or deep learning techniques.[63] Among those, 25 articles were on Natural Language Processing (NLP), encompassing a diverse array of topics such as English to Korean machine translation, text to speech algorithms, keyword extraction, database construction, and detection of banned words from text. Thirteen articles were on Computer Vision, encompassing topics such as object detection, image segmentation, and image enhancement. Six articles were on autonomous vehicle technology and three articles were published on methodology. These indicate that there is significant effort at Kim Il Sung University to invest in AI research for a wide range of applications.

Most importantly, one 2018 *Gazette* article specifically dealt with machine learning applications for cyber defense, by illustrating feature section methods for improving network intrusion detection rates.[64] This article used publicly available training data from the 1999 Knowledge Discovery and Data-Mining (KDD) intrusion detection competition run by DARPA and MIT Lincoln Labs. In addition, while not strictly using machine learning or AI techniques, three other articles on enhancing cyber defense was published, two on methods to prevent Denial of Service (DoS) attacks and one on intrusion detection for industrial control systems.[65] Perhaps unsurprisingly, the journal did not publish any research on offensive cyber operations or using machine learning or AI to enhance attacks. At a minimum, these articles indicate that there are efforts to improve North Korea's cyber defense and nascent attempts to do so using machine learning techniques. The fact that no such articles were published by the Gazette in 2020 is hardly evidence that such research has ceased. The absence of such articles is more likely to be the result of North Korea's desire to keep such research secret. Given North Korea's emphasis on offensive cyber operations, it is particularly likely that similar research is being conducted on a more carefully concealed basis, to leverage such techniques to enhance offensive cyber operations as well.

We assess that it is likely that similar research is being conducted in North Korea's other universities and research institutes, such as Kim Chaek University of Technology, Pyongyang University of Science and Technology (PUST), and the Institute of Information Science and Technology at the State Academy of Sciences. Furthermore, at least one business venture—Yalu River Technology Development Association—is reportedly developing various AI-enabled biometric identification systems including facial and speech recognition software.[66]

North Korea also seems to be conducting research on the computing infrastructure needed for AI. The Institute of Information Science and Technology and Kim Chaek University are both reportedly conducting research on cloud computing services in North Korea and have published articles on the topic.[67] North Korea has also been conducting research on parallel computing in order to increase computing power using CPUs without relying on scarce GPUs or TPUs.[68] In fact, a Kim Il Sung University article published in 2020 presented one method of using a convolutional neural network (CNN) to identify license plates quickly in computers equipped with only CPUs rather than GPUs.[69] Such research indicates that while North Korea certainly faces material constraints in computing, these have not stopped its researchers from continuing to pursue research on AI topics.

There is also evidence that North Korean researchers are trying to actively utilize open-source datasets and models for their research, and that the North Korean government is prioritizing such efforts despite economic hardship. For example, a 2021 research paper published in the math series of Kim Il Sung University Journal uses deep learning to make diagnoses based on chest X-ray image data.[70] This paper uses a relatively recently published 2019 dataset called MIMIC-CXR Database v2.0.0, suggesting that researchers are keeping up to date on the latest academic research outside of North Korea.[71] North Korean researchers have also used GoogLeNet for video and image analysis in the areas of license plate identification and facial recognition.[72]

Despite their ability to access open source writings from other countries on AI, a number of other factors, including access to human talent, data, electric power, computer hardware or other resource restrictions could constrain the North's development of AI. These are discussed in the next section.

Constraints on North Korean AI Development

This section addresses North Korea's requirements for achieving an AI-enabled cyber capability and its potential paths toward this goal. Broadly speaking, the core building blocks of AI-enabled cyber capabilities include computing hardware and software, large-scale coded databases on which to train the AI programs, the electricity to run the programs, and the human capital capable of organizing and drawing up advanced algorithms to surface patterns in the data. Overall, we assess that North Korea's human capital limitations and challenges in accessing large-scale advanced computing are likely the biggest challenges for the regime's AI efforts, though these are unlikely to prove insuperable and can likely be mitigated in several ways should the regime decide to allocate resources to doing so. Although North Korean authorities are likely masking more sensitive AI research (particularly those related to offensive and defensive cyber operations) from public media, the available evidence suggests that North Korea's general research on AI remains at a fairly early stage of development, but will likely continue to advance.

Countries have a variety of pathways to adopting and employing emerging technologies.[73] If truly closed to the outside world, they can develop a technology indigenously, or if they can leverage the global community, they can buy it outright, copy, or steal it from abroad, or cooperate with others by sending researchers to train in other countries. In the case of North Korea, one possible approach that can shed light on the challenges the regime faces is to compare its efforts to acquire AI to its acquisition strategy for other strategic technologies in which it has an interest, including nuclear weapons, ballistic missiles, and cryptocurrency (useful for money laundering and sanctions evasion).[74]

North Korea's nuclear program was founded, and its Yongbyon nuclear complex built, with assistance from the Soviet Union and some from China, but grew into a strong indigenous program that was informed by outside expertise. After the end of the Cold War and Pyongyang's withdrawal from the Non-Proliferation Treaty brought widespread international opposition to North Korea's nuclear program, Pakistani scientist AQ Khan's international network proved to be a key source of assistance for developing North Korea's clandestine uranium enrichment program starting in the 1990s.[75] Since then, it appears that Pyongyang further developed its nuclear know-how largely on its own, based on a foundation of domestic education and human capital, though the growing body of openly available information on nuclear technology meant that North Korea could still draw on foreign information sources to fill knowledge gaps. Similarly, North Korea's cyber and AI programs appear to have both a strong foundation of indigenous knowledge and human capital to build upon, along with access to relevant foreign information sources.

In theory, the computer hardware required to develop AI should not be easy for North Korea to obtain. Such electronics are considered banned for export to North Korea by the United States and many other countries, under the provisions of UN sanctions in the categories prohibiting imports of "luxury goods" and "dual-use" technologies, but China and other countries do not share such an interpretation of sanctions resolutions. Broad U.S. export controls also prohibit trade of such items with North Korea, but North Korea has been able to obtain and rely upon technology from U.S. companies for its cyber programs.[76] To obtain hardware in violation of UN sanctions, the North has relied on an ever-evolving global network of front companies, North Korean and foreign proxies operating in a number of countries to acquire discrete hardware components.[77] These transactions rely on sellers either ignorant of the true DPRK end-use or complicit in such sales, and generally evade notice by relevant authorities.

In obtaining technological know-how for technologies such as blockchain and cryptocurrency, the regime has been able to woo a handful of foreign specialists to advise it by positioning itself as a source of funds and a voice for enthusiasts of "disruptive technologies."[78] One such foreign expert was US citizen Virgil Griffith, who recently pled guilty to providing cryptocurrency services to North Korea and assisting it in sanctions evasion.[79] A regime move to invite foreign AI/ML experts to advise it similarly could be an early indicator and warning that the North is moving towards a more fulsome embrace of AI/ML for cyber.[80]

North Korea is less likely to face major obstacles to acquiring AI technologies than those undergirding nuclear weapons and ballistic missiles, since the latter are far more internationally controlled than the information associated with AI and ML. As another expert we spoke with commented, "North Korea is very quick at adopting ideas and technology from outside, and this is much easier to do with respect to cyber and AI/ML code than it is in the nuclear hardware space."[81] Pyongyang is likely to seek out AI-related hardware and software from other countries through a variety of avenues and, despite its ideology of *juche,* or autarky, will gladly buy AI-related capabilities from abroad when it can find the right seller and the right price, including through the use of front companies. While the Kim regime has sufficient capital to fund key regime priorities, such as its nuclear weapons program and supporting its elite lifestyle, it is unclear how much money it will dedicate to AI-related technologies, but it can almost certainly afford to dedicate enough to develop key capabilities. Indeed, despite predictions by some, including the Korea Development Bank's Future Strategy Institute, that international sanctions would cause the North's AI development to "hit a wall," the regime's AI technologies appears to be continuing to make progress, even if slower than would be the case were sanctions lifted.[82]

When it is unable buy the necessary technologies outright, North Korea will likely either try to copy (replicate) or steal it from ripe targets.[83] Stealing AI technologies is possible, most likely through cyber theft, though there is no guarantee North Korea would be able to actually use these technologies to their fullest extent.

Another option, sending DPRK nationals abroad for training, appears to be a route that North Korea is pursuing for AI. For one sense of scale, a 2016 report from Voice of America stated that roughly 50 to 60 North Korean students with backgrounds in science, technology, engineering, and mathematics (STEM) are selected annually to study abroad as a part of the regime's initiatives to develop its cyber capabilities.[84] While North Korea appears to have sent its researchers out to several countries, including Russia, Italy, India, and Romania, among others, China appears to be the main source of foreign training for DPRK researchers working on strategic technologies, and is thus the country of greatest concern.[85] China is known to loosely control what North Korean students study at Chinese universities, and may have violated UN sanctions by allowing students from the DPRK to study a range of dual-use technical subjects even after stricter sanctions were imposed in 2016.[86] For our part, we were able to identify at least one DPRK researcher, Kim Chungsong from Kim Chaek University of Technology's School of Automation, who completed a Ph.D. on deep learning at China's Harbin Institute of Technology in 2019.[87]

One key distinction between nuclear weapons and AI technology, however, is that there is a wide swath of open-source technology and information available that anyone can download as it is a nominally civilian and largely unregulated set of technologies. The amount of publicly available information on AI dwarfs the amount of detailed information publicly available on nuclear issues. For example, publicly available and free open-source data sets for machine learning means that North Korea does not in all cases have to create its own training data.[88] Moreover, the availability of Tensorflow and similar machine learning platforms means that North Korea does not even have to create its own algorithms, though it may need to modify these for its specific purposes. The degree of scrutiny on end-users by major providers of key services, such as Amazon Web Services (AWS), is unclear but likely substantially less than on nuclear or missile technologies. Nonetheless, the challenge for the North will be to gain access to good, representative, and reliable data on which to train its algorithms, and the North would surely face some, though probably not insuperable, obstacles to acquire it. They may need to steal one or many datasets, ensure that as these are merged they are all coded to the same standard, and make certain that their data fits their AI model. Ensuring this could be a major complicating factor for North Korean employment of AI for offensive cyber.

As it seeks to access these key components of AI, North Korea is most likely to work, at least in part, through its neighbor and patron, the People's Republic of China (PRC). China's position as Pyongyang's biggest source of high-tech products and cutting-edge scientific knowledge owes in substantial measure to the lack of commitment by the PRC to enforcing existing UN sanctions. As the UN Panel of Experts 2019 report found, a DPRK Academy of National Defense Science-affiliated company signed a 2019 agreement with a Chinese company to employ three North Korean programmers and two hardware developers, to design "artificial intelligence products, including both software and hardware."[89] While this appears to be an instance where the DPRK was providing AI services to a Chinese company, it nevertheless shows the ability for DPRK AI experts to flow relatively freely between the two countries, and may be an example of the Kim regime finding opportunities to export its AI specialists abroad where they can learn additional skills.

North Korea may also seek to exploit foreign experts to speed its AI development, a pathway it appears to have already attempted to leverage in its separate efforts to understand how blockchain and cryptocurrency could be used to circumvent international financial sanctions.[90] Unlike the practical requirement for much of nuclear weapons development that it be done in-person and thus would require willing foreign experts to travel to North Korea, cyber technologies can be developed and employed from almost anywhere in the world. This enables North Korea to either hire a complicit foreign expert and have them work remotely, or even hire foreign experts via cutout or front companies.

As to whether or not consistent electricity flows could pose a constraint on North Korea's development of AI, while the DPRK is generally regarded as an energy-poor country, there are almost certainly sufficient electricity resources for the regime's AI advancement to make progress if the Kim regime decides to prioritize it. Clearly the much more energy-intensive nuclear and missile programs the DPRK has pursued have not been substantially constrained by access to a consistent power supply. Indeed, North Korea actually exported almost 320,000 megawatt hours (MWh) of electricity to China in 2017, worth an estimated $11 million, and could repurpose some of that electricity should any AI development programs face shortfalls.[91]

In terms of powering a machine learning data center, estimates depend on the computing power, the hardware's efficiency, and other power demands (cooling, etc.). One estimate for a generic facility in 2019 put it at 52.8 MWh/day, and reports suggest one particularly large Chinese bitcoin mining operation in Iran was using 175 MWh/day in early 2021.[92] For comparison, a $900 million U.S. National Security Agency supercomputing center built in the early 2010s was estimated to require 60 MWh whereas training a single natural language processing (NLP) model from 2019 could require 635 MWh (over multiple days).[93] This suggests that North Korea's electricity exports to China, if repurposed, would likely be more than sufficient to cover its AI-related power demands.

Access to the Internet is not likely to prove an insuperable stumbling block for Pyongyang either. Despite North Korea's domestic limits on Internet connectivity, key regime organs clearly have access inside the country as necessary through connections via China and Russia.[94] Moreover, North Korea is generally believed to station at least some of its cyber operators overseas, both for deniability and better access, including in China, Russia, the Philippines, Malaysia, Cambodia, Belarus, and India, among others.[95] However, the 2019 UN Panel of Experts report stated that "nearly all malicious cyber activity for the DPRK now comes from inside the DPRK itself and is conducted by the malicious cyber actors themselves," while the IT workers the DPRK sends overseas are largely directed towards generating revenue for the regime, meaning Pyongyang bans them from hacking so as to avoid attention.[96] Either way, the Kim regime is clearly able to access the Internet. Perhaps the only challenge would be Internet speed if the regime wanted to leverage foreign cloud computing providers.[97]

While data, expertise, sufficient electricity, and Internet access likely would not pose insuperable obstacles, computer processing capacity could pose a somewhat greater challenges for the regime's efforts to advance its AI capabilities. As one assessment has noted, a major problem for the regime could be the challenges associated with "the acquisition of equipment that cannot be produced indigenously… and the high-costs of high-performance computing and other equipment."[98]

While North Korea clearly has no difficulties in procuring some high-end computers—Kim Jong Un has been spotted with several Apple products, for example—the sheer scale required for advanced AI may be more of a hurdle.[99]

Individual machine learning workstations can be built for $5,000 or less to process data locally, or the DPRK could try to leverage cloud computing.[100] Cloud computing can be a relatively cheap way to process data, but becomes expensive the larger the dataset is and the higher the accuracy desired. Beyond AWS and similar Western counterparts, Chinese technology heavyweights such as Alibaba, Tencent, and Baidu all have cloud offerings. PRC cloud computing suppliers are likely both cheaper and less likely to be scrutinized rigorously by actors looking to prevent North Korea from advancing its AI capabilities. Indeed, one expert interviewed by the authors stated that there is evidence that DPRK cyber actors are already using Chinese cloud systems, though likely not for AI-related activity.[101] On the other hand, if secrecy or cost are priorities, cloud computing may prove an unlikely path for Pyongyang to exploit as it advances its AI capabilities. Overall, however, at present most of North Korea's computing capabilities are limited to older and less efficient systems running on less advanced computational hardware, which could slow its AI development somewhat, especially if AI development has to compete with other technology areas that also demand access to advanced computing. The North's AI development is also limited by its reliance on central processing units (CPU), as opposed to the graphics processing units (GPU) that are more commonly utilized in AI development.[102]

An analysis of the factors that go into AI/ML development can suggest areas that may prove to be constraints on the North's development of these technologies. Assessing how other countries have developed their own AI/ML systems, including for cyber, can shed additional light on possible pathways the North might seek to travel, either because these are universal or because Pyongyang seeks to learn from their experiences.

Comparisons with Other Countries

Given the challenges of studying North Korea's thinking about AI and cyber directly, what insights might be gleaned from comparison with other countries? And how does North Korea view other nations' efforts on AI and cyber? Some countries, such as Iran, have long engaged in official cooperation on national security technology affairs with North Korea, whereas others, such as China and Russia, have more tacitly allowed Pyongyang to benefit from their knowledge, experience, and expertise. For its part, the North strives to learn from "everyone," and "they study their partners and their enemies alike."[103] Even though the DPRK does not have formal diplomatic relations or scientific exchanges with countries like the United States, South Korea, or Japan, its researchers will "read as many [articles in] academic journals about foreign artificial intelligence and machine learning [as applied to cyber] as possible," and will likely be able to learn a lot from this since "the science, math, and engineering [underpinning these fields] are all openly distributed."[104] The North has also allegedly targeted security researchers on Twitter, LinkedIn and other social media platforms with job offers or proposals to collaborate on vulnerability research.[105]

As experts on North Korea we spoke with argued, Pyongyang will also study leading nations in the field out of a desire to understand how its adversaries think about such technologies and in order to uncover any vulnerabilities that such research may reveal.[106] One important caveat is that all these countries are richer and more connected to the Internet than North Korea, and paradoxically, this may give the DPRK a comparative advantage in one respect: its lack of connectivity to the Internet means it has fewer gateways to monitor and consequently needs to expend fewer resources on cyber defense.

Under Xi Jinping's leadership, China has made developing AI a key national priority, including for military applications.[107] While the PRC has invested in its own AI capabilities, it also benefits from knowledge exchanges with the United States and others.[108] According to a CSET estimate of Chinese AI spending, Beijing is investing "a few billion dollars [per year]" on AI, levels roughly similar to that of the United States.[109] Thus a May 2020 indictment of DPRK cyberhackers, which claimed they had stolen US$2.5 billion, suggests the regime likely has sufficient funds available to support development of AI if it prioritizes such.[110] The most relevant point may simply be the fact that China is North Korea's neighbor and ally, as well as an important model for the regime's sense of what is technologically feasible. If North Korea obtains even limited access to AI technologies from China, this could be a substantial boost to DPRK AI capabilities, including potentially its cyber toolkit. Moreover, North Korea is certainly watching China's growing adoption of disinformation, especially on social media (with the PRC's activities increasingly influenced by observation of Russian disinformation), and could take inspiration from this to pair its advancing AI capabilities with efforts to target foreign societies or individuals with increasingly believable (if deceptive) messaging.[111] In doing so, Pyongyang would likely simply be building on and adapting its already extant efforts to leverage nationalistic propaganda on social media, while layering on false or distorted narratives supported by fake audio or video content that could be made more convincing through AI augmentation.[112]

Russia may represent another useful comparison for how North Korea could move toward more capable cyber capabilities if paired with AI. In 2017, Vladimir Putin said that "artificial intelligence is the future, not only for Russia, but for all humankind… Whoever becomes the leader in this sphere will become the ruler of the world."[113] Of the three global actors surveyed here, perhaps the most is known about Russia's offensive cyber capabilities, including their shortcomings. Moscow has probably employed the most destructive offensive cyberattacks of any actor, including the NotPetya attack on Ukrainian critical infrastructure that

spread around the globe.[114] This was similar to North Korea's WannaCry virus, but more dangerous. One recent study argued that Moscow's biggest challenges for improving its cyber capabilities, including adopting emerging technologies such as AI, are talent recruitment, personnel retainment, and bureaucratic dynamics.[115] For North Korea, recruitment and retention should generally be easier because of the regime's ability to exercise strict social controls on education and labor allocation, though bureaucratic dynamics exist within any government and could be an issue. Russian analysts themselves have expressed skepticism that Russia can keep up with the United States and China in its adoption of AI, though they note Russia could still be a "serious player" and "local leader."[116] Outside traditional cyber operations, Russia's potential application of AI for cyber-enabled political warfare could also be an attractive model for North Korea, which, like Russia, also employs online disinformation campaigns against its enemies.[117]

In some ways, Iran might be the most useful comparative case due to its successful development of cyber tools despite international isolation that, while not as great as North Korea's, is nonetheless far greater than that of either China or Russia. Separately, although Tehran does not have a national AI plan, it has nonetheless pursued AI as well as other emerging technologies with zeal.[118] It has also carried out destructive cyber-attacks against Saudi critical infrastructure and other industrial control systems.[119] Of possible relevance, while Iran's isolation may have slowed the Islamic Republic's technological development, it has not stopped it from developing increasingly capable cyber tools and building advanced supercomputers, foundational building blocks of any future AI-enabled cyber capabilities.[120]

In addition to the aforementioned countries, our interviewees were careful to make sure to point out that North Korea also seeks to learn from the experiences and expertise of the United States, South Korea, Japan, and other advanced AI and cyber powers. The United States Department of Defense, for example, has reportedly initiated research on how AI might be employed to accelerate cyberattacks and improve cyber detection and defense.[121] In 2016, partly fueled by North Korean cyberattacks, South Korea and the U.S. jointly launched an effort to develop AI-based cybersecurity tools.[122] And Japan's Ministry of Defense announced a budget of nearly US$240 million in 2020 to bolster machine-learning for cyberdefense.[123] Given that much knowledge about artificial intelligence circulates openly in Western scholarly journals, North Korean researchers can likely draw on this without leveraging espionage or criminal activity.

POSSIBLE WAYS NORTH KOREA MIGHT PAIR AI/ML WITH CYBER

This section provides an assessment of how North Korea may leverage AI/ML techniques to improve its cyber capabilities generally. Over the past decade, North Korea's approach to cyber has been primarily focused on exploiting adversary vulnerabilities for political and economic gain, rather than using the Internet to increase productivity at home. North Korea's approach to AI may follow a similar trajectory, seeking to improve its cyber operations. A successful cyber operation often relies on exploiting a defender's vulnerability, and therefore resembles a perpetual cat-and-mouse game between the defender's efforts to discover and mitigate vulnerabilities and the attacker's efforts to exploit them, though this may shift as newer systems come online with more proactive security approaches.[124] In recent years, AI has emerged as a potentially important tool in boosting the effectiveness of both sides of the equation. Given North Korea's predominant focus on cyber offense rather than defense, we assess that it will likely prioritize AI/ML for cyber offense, though perhaps while focusing even more on automation, at least in the early stages.[125]

The phrase "cyber offense" as used here encompasses any intrusion into foreign networks and systems for the purposes of espionage, attack, or criminal activity. This entails a broader range of activities than the concept of "offensive cyberspace operations" as used in the U.S. military.[126] Table 2 lists some of the most common types of cyberattacks attributed to North Korea.

Table 2. Common Types of Cyberattacks North Korea Has Employed

Tactic	Description
Phishing/Spear Phishing	Practice of sending fraudulent communications that appear to come from known, reputable sources. Phishing attacks are distributed widely without attempting to target a specific victim; spear phishing attacks are more narrowly tailored attempts to target a specific individual.
Traffic Generation	Employed in distributed denial of service (DDoS) attacks, this occurs when the attacker floods networks and servers with excess traffic, leading the system to become incapable of fulfilling legitimate requests.
Command/Code Injection	An attack that inserts malicious commands or code that are then executed by a system; one example is a Structured Query Language (SQL) injection, which can result in Information Disclosure.
Zero-Day Exploit	Attacks a previously unrecognized vulnerability in a software product before a security patch or solution has been implemented.[127]

Tactic	Description
Ransomware Attack	Type of malware attack that utilizes encryption to hold victim's information hostage; access to the information is only restored after the hacker's demands are met.
Watering Hole Attack	Exploitation method where hacker seeks to compromise a commonly visited website with the aim of infecting the computers of all those who visit it, using such infections to gain access to the victims' networks.
Data Exfiltration	The stealthy removal of information from a network, used in Information Disclosure attacks. One common approach is Domain Name System (DNS) Tunneling, which utilizes DNS protocol to communicate non-DNS traffic and can be used to conceal outbound traffic past typical network defenses.
Remote Code Execution	The ability to execute arbitrary commands on a target system. Remote execution can be enabled by malware or through the presence of a vulnerability that permits code injection.
Credential Harvesting	Gathering of usernames, passwords, encryption keys, and other information that can be used to access systems or resources.

Sources: CISCO, McAfee, Kaspersky[128]

Aside from the feasibility of leveraging AI for cyber offense, there remains a more fundamental question as to whether North Korea will believe that investing in such sophisticated techniques confers a marginal benefit. On the one hand, many targets of North Korean interest already have vulnerabilities that can be exploited with existing tools, and as noted above in some cases North Korea has simply purchased access to targets from criminal groups.[129] There are already existing tools that can increase the destructiveness of an operation without AI/ML techniques, either by using worms or by targeting software supply chains.[130] Extensive reconnaissance can already be conducted on a target through open sources and credentials and personal information can often be purchased on the Dark Web. The majority of North Korean cyber operations are likely to continue in this fashion.

On the other hand, as more and more organizations rely on AI/ML-based cyber defense systems from security vendors, investing in adversarial machine learning is likely to present new opportunities for North Korea. This will especially be the case if organizations harbor a false sense of security after adopting AI-based defenses, giving North Korea the element of surprise on a critical mission.[131] Another potential benefit to North Korea of leveraging AI for cyber offense is that it could enhance the regime's ability to conduct operations, which is currently constrained by the relatively limited number of high-quality hackers it can train each year from its domestic talent pool. Given the North's public obsession with "self-reliance," the potential to increase throughput of either its cybercrime or espionage activities without relying on third party criminal groups or foreign assistance may be an importance source of motivation to continue research in this area. Another possibility is that North Korea may simply decide to invest in such research early on even though they may not have immediate plans to utilize them in actual cyber campaigns, given their general interest in AI research.

Evasion or Exploratory Attacks on AI-enabled Cyber Defenses

An important element of cyber defense rests on detecting malicious activity by classifying a file, text, or activity as normal or abnormal, for example in the areas of spam, intrusion, or malware detection.[132] In recent years, advances in AI and ML research, as well as the increasing availability of computing resources, have shown promise in certain specific contexts for improving the accuracy and efficiency of such tasks. In particular, unsupervised learning and deep learning techniques can help identify novel threats, complementing the existing approach of comparing new, incoming activity against a database of known, labelled normal or abnormal behavior, though the overall contribution that such approaches can make is heavily context-dependent.

Some of the most common forms of cyber defense mechanisms utilizing AI include automated detection technology, such as antivirus software that utilizes machine learning to identify and respond to malware attacks. Further examples could include expanded automation and increasing the generalizability of automated intrusion, anomaly, and malware detection software, network analysis, and spam filters that look for emails carrying malicious code. AI can also be utilized to automate security vulnerability scans of networks and systems which can help to improve network security. These types of automated scans allow for cybersecurity professionals to monitor for attempted intrusions, detect anomalous activity, do keyword matching, and track various statistics that can help to prevent and/or accelerate response to cyberattacks. Additionally, automated anti-phishing tools can effectively scan emails and other online communications for malicious attachments, links, and message bodies before they reach the user.[133] These types of automated cyber defenses may be helpful and are increasingly being adopted across the globe to deal with growing levels of automated cyberattacks.[134]

However, the adoption of such tools can be a double-edged sword. AI- and ML-enabled detection can present a new set of vulnerabilities, and an attacker may exploit these in order to evade a target's cyber defenses. An AI-based detection system is potentially subject to a variety of classes of attacks.[135] First, adversarial examples can be created in an *evasion attack* to cause misclassification of a particular input. Second, false data can be added to the training data in a *poisoning attack* to affect the model itself, systematically misclassifying inputs or decreasing confidence levels. Third, an *exploratory attack*, such as a membership inference or model inversion attack can jeopardize privacy by gaining information on aspects of the underlying training data used in the model, or to create a model that performs similarly to the original one. Each of these can be either white box attacks, in which the attacker has full knowledge of the model and the underlying training data, or black box attacks, in which the attacker has no knowledge of the model or the training data, and can observe only outputs from an input, often by querying an application programming interface (API), or the set of software rules, definitions and protocols that enable two or more applications to communicate with each other.[136]

Cybersecurity researchers have found evasion attacks to be probably the most common among the three types of attacks, often undertaken during the inference phase.[137] Several academic papers have successfully generated adversarial examples that evade AI-based malware detection models. For example, a number of research efforts have demonstrated that gradient-based attacks can be used to form examples that cause misclassification in MalConv, a malware detection model using a Convolutional Neural Network (CNN) trained on raw bytes.[138] Even worse, and potentially more in keeping with the DPRK's overall approach to cyberattacks, might be the development of multiple attacks that generate false positives that can overwhelm or discredit a system.

In addition, security researchers have found that many AI-based models already being used by customers exhibit substantial fragility. In 2019, security firm Skylight found that by adding a few pieces of benign code at the end of a malware file, they were able to cause Cylance's AI-based malware detection model to misclassify the malware as benign.[139] By doing so, they were able to cause misclassification in almost 84 percent of the 384 samples tested, including well-known malware such as Mimikatz.[140] In 2021, Kaspersky researchers conducted a variety of white box and grey box attacks against their own ML-based malware detection model, and found that they were able to frequently cause misclassification, including 89 percent of the files tested for a particular white box attack.[141] Although generating adversarial examples for texts is assumed to be more difficult, more research is also being conducted on evading spam filters by using NLP.[142] These researchers' findings show that AI-based cyber defense models are far from silver bullet-type security solutions and are themselves vulnerable to a variety of attacks, and sometimes suffer from a lack of explainability.

Over the next decade, North Korea may try to apply academic AI research to adversarial machine learning. If they do so, the DPRK's hackers may be able to find ways to make small perturbations to their malware without jeopardizing its overall functionality, ultimately evading some of the AI-based defenses used by their targets. As more potential targets deploy AI-based cyber defenses, and as relevant academic research develops, the prospects of cyberattacks leveraging adversarial examples will likely increase.

Would AI-Enabled Cyber Capabilities Change DPRK Behavior?

The scope of such potential future AI-enabled cyberattacks could be narrow or wide, either against a particular target that happens to use an AI-based malware detection system, or indiscriminately against any targets that rely on an AI-based cyber defense system. In the former case, North Korea would likely have an incentive to exercise restraint, similar to Russia's approach when it chose not to exploit the vast majority of the potential victims it had access to by virtue of their downloading the compromised SolarWinds Orion software. In such intrusions, widespread exploitation of such illicit access increases the chance that an intruder will be detected and expelled from the network before completing its campaign against the real target.[143] Indeed, North Korean hackers have already shown that, like their Russian counterparts, they too can be patient in pursuit of longer-term gains. In 2016, South Korean police alleged that Pyongyang had hacked "more than 140,000 computers at 160 South Korean firms and government agencies, planting malicious code under a long-term plan laying groundwork for a massive cyberattack against [the ROK]."[144]

Alternatively, Pyongyang may take a more opportunistic approach, especially if the goal of a given cyber operation is financial gain rather than espionage or support to an ongoing military campaign. In such a case, North Korea's behavior might be similar to the ransomware attack in which the REvil gang indiscriminately exploited large numbers of compromised Kaseya software users.[145] As the foregoing suggests, AI-based defenses may present a software supply chain risk for companies and individuals relying on the product, turning AI from a tool for cyber offense or defense into a target in its own right. In this case, North Korean hackers might try to make the most of the window of opportunity before the vulnerability is mitigated, similar to how China installed backdoors on almost every Microsoft Exchange server on the Internet before vulnerabilities were patched.[146]

North Korea may also use adversarial machine learning for *exploratory attacks*, in which attacks are designed to learn more about the model itself or recover aspects of the underlying training data. These attacks pose significant privacy concerns as models are often trained using sensitive data meant to be kept confidential. North Korea is likely to use the information

extracted from these attacks to make advances in its own AI research, and to fine-tune its adversarial examples for more effective evasion of cyber defenses. Although it may be feasible for North Korea to use the recovered personal data as reconnaissance for further cyber operations or as blackmail for extortion and revenue generation, the likelihood is low as existing tools can already achieve such effects with less effort.

The confidentiality of a model and its underlying training data can be compromised in a variety of ways. One way is through membership inference attacks, which try to identify whether a particular data point belonged to the training set used to train a model.[147] Identifying that a particular individual belonged to a chest X-ray training set, for example, could reveal that individual was a patient at a particular hospital that participated in such data collection efforts. Another type, called a property inference attack, aims to learn global characteristics about the training data such as the testing environment or the distribution of the dataset rather than any particular data point.[148] A third type, the model inversion attack, aims to retrieve the underlying training data by querying models. For example, one study showed that it was possible to recover how an individual answered to a survey asking about marital infidelity, and also reconstruct facial images by querying a facial recognition service and a unique identifier such as the individual's name.[149] Finally, a model extraction attack aims to "steal" the model by replicating the functionality of the model without prior knowledge of the model's parameters or training data.[150] Researchers have even shown that a model extraction attack and a model inversion attack can be combined to improve the efficiency of attacks under black box settings.[151]

These attacks are relevant for models using cloud-based Machine Learning as a Service (MLaaS) on platforms such as BigML and similar services offered by Amazon, Microsoft, and Google. Users of such services, such as a facial recognition service, often make the resulting model publicly available for prediction queries via an API, allowing others to repeatedly interact with the model to obtain outputs. Both white box and black box attacks are possible against these models with varying efficiency.

North Korea may find such attacks useful for specific tasks where they face significant difficulties obtaining quality training data from domestic or open sources alone, which in turn would hinder their ability to conduct ML and AI research in that area. Although recovering an entire dataset with high accuracy would be a difficult task, North Korea could conceivably use a combination of model extraction and model inversion to create an approximate version of the model, without having to gather the data or train the model from scratch.

In particular, North Korea may try to conduct exploratory attacks on AI/ML classifiers pertaining to cybersecurity, in order to further aid evasion attacks. Researchers at the University of Louisville, for example, have trained classifiers on datasets including Spambase (spam detection), KDD99 (intrusion detection), and CAPTCHA (bot detection), and presented an algorithm for reverse engineering the model and generating adversarial examples that evade detection.[152] Similarly, analysts have conducted an experiment extracting a classifier using the KDD99 dataset.[153] And computer science scholars in France and the United States have extracted a variety of models available in BigML, including a model predicting whether an email will be classified as spam or not, given the message content.[154]

For their part, in 2018, North Korean researchers at Kim Il Sung University conducted their own research on improving intrusion detection using the KDD99 dataset, and their exploration of this topic may be just a matter of connecting the dots, given their interest in this area.[155] Models and training data extracted from such exploratory attacks can help North Korea make its own cyber defense research more robust, and could also be used to generate adversarial examples that more effectively evade defenses. Such efforts could help North Korea's hackers stay competitive despite their relative lack of good training data pertaining to cyber activity.

Finally, exploratory attack methods such as membership inference pose significant privacy risks of revealing sensitive information about the training data. North Korea is probably less likely to use such methods for the purposes of gathering intelligence about a target before conducting a cyber operation or for blackmailing a target, since there are far easier ways to achieve the same goal. Whether by using conventional cyber exploitation or even simply by searching open sources such as a target's social media profile, adversarial learning offers little marginal benefit compared to the labor and computational resources required to achieve the task.

Other potential applications need more careful evaluation, such as using NLP techniques to enhance the scale and effectiveness of spear phishing campaigns.[156] On the one hand, it is possible that North Korea may try to create more convincing English-language texts through using models such as GPT-J and GPT-Neo, which are open source alternatives to the Generative Pre-Trained Transformer-3, a program that uses deep learning techniques to generate realistic written material.[157] North Korea could also attempt to build more convincing fake social media personas though the use of generative adversarial network (GAN) programs to create profile pictures and NLP-enabled chatbots, for future campaigns similar to the one in which North Korea created several fake Twitter and LinkedIn profiles to steal vulnerabilities from security researchers.[158] As noted above, North Korea is currently conducting extensive research in NLP, including efforts to build a North Korean text and speech corpus, as well as research in keyword extraction and machine translation.[159] If such research makes significant progress, it may be possible to increase the scale of phishing campaigns without having more English-proficient hackers, and also increase the effectiveness of social engineering by being able to tailor the content specific to the target in question.

On the other hand, it is also important to consider the actual likelihood that North Korea will apply such sophisticated techniques for spear phishing campaigns. Often, North Korea can simply obtain initial access to targets by purchasing them from criminal groups or getting insiders to click on malicious links via email attachments or watering hole attacks. Techniques such as those above, if they are to be ever used, are more likely to be seen against high-value targets that would be difficult to access otherwise.[160] Other cutting-edge academic research is being conducted in areas such as generalizable vulnerability discovery algorithms that could have robust cross-architecture applicability; however, research in these areas is still nascent, and it is too early at this point to arrive at any firm conclusions about its implications for North Korea.[161]

As countries such as the United States, the Republic of Korea, and Japan move toward greater adoption of AI for military operations, North Korea's interest in exploring how it could use cyber to attack AI systems is likely to grow. As a 2018 RAND report noted, "many experts were concerned by the potential for an adversary to subvert even a very capable AI by hacking, poisoning its training data, or manipulating its inputs."[162] At the same time, the demonstration effect of other countries' military AI advances will likely incentivize the North to keep pushing forward in this area. Most notable in that respect are likely to be China's extensive writings on the "intelligentization" of modern warfare, and its establishment of an integrated Strategic Support Force focused on information warfare.[163]

CONCLUSION

For now, North Korean offensive cyber activities appear to be reliant on human operators who design and execute the wide range of cyber operations that the regime has perpetrated. There is similarly little evidence that the regime has empowered advanced AI to play a key role in the defense of its network systems, though it likely has relatively widespread adoption of automation built into its firewall, network monitoring, censorship and keyword filtering software. We were also unable to find much evidence that North Korea has yet sought to attack AI systems' training data or attempted to poison the model that such systems use to make decisions in support of network defenses. It is unclear whether this is because of limitations in adoption of attacks on AI systems' training data sets globally, or because North Korea simply has not yet undertaken such efforts to target adversary AI training sets, or because we simply do not know about North Korea's cyber activities against such systems.

Yet, despite the absence of evidence that the regime has adopted AI for cyber at this point in time, we do find substantial support for concluding that the DPRK is interested in AI, is tracking international developments in the field, and is investing in its own capabilities, as demonstrated by leadership comments, official media reports, and scholarly research outputs from leading North Korean universities, as well as discussions with subject-matter experts. These factors, combined with the regime's clear and growing emphasis on cyber capabilities as a part of its national security toolkit, strongly suggest that the regime may make the leap to AI-enabled cyberattacks, cyber defense, or targeting of adversary AI systems using cyber means in the near future. If it does, we assess that the regime's motives, targets, example and the impact AI could have on its cyber operations could look like Table 3.

Table 3. Possible North Korean Uses of AI-Enabled Cyber for Regime Goals

Kim Regime Goals	How Cyber Contributes to this Objective	How Might AI Be Incorporated
Regime Security	Raise money for regime	May already leverage large-scale, automated botnet construction to improve brute force, scale, and speed of ransomware and banking attacks
	Protect regime "honor"	Use of modeling attack to gain access to, "punish" adversary through tailored leaks of private data
	Defend regime networks	Expand use of automated intrusion detection and malware defense
Breaking U.S.-ROK Alliance	Drive political-military wedges in U.S.-ROK alliance	Create large numbers of high-quality AI-tailored messages appearing to come from trusted sources alleging criminal activity by USFK forces or claiming U.S. driving up risk of war
	Collect military intelligence	Faster, more frequent cyberattacks on military/intel targets
	Undermine military capabilities	Accelerate pace of lateral transfer within systems once access achieved

Kim Regime Goals	How Cyber Contributes to this Objective	How Might AI Be Incorporated
Unifying/Dominating Peninsula	Collect political intelligence	Collect, sift, bin and weaponize data on ROK elites, public opinion through AI-enabled targeting and cyber data theft
	Undermine support for ROK government	Seek to drive public opinion in ROK through improved disinformation, texts, emails, images, and sound to drive down support for politicians who favor resisting, and build support for those who embrace, North Korea
	Asymmetrical coercion	Execute expanded harassment and intimidation of NK adversaries through improved stealth and reduced detectability/attribution of attacks

Moreover, our review of possible constraints on the DPRK's ability to develop AI for cyber suggest that cost, access to expertise, or electricity supply are not likely to prove insuperable stumbling blocks, though constrained access to high quality coded data for training sets, to human talent, and to sufficiently advanced high-powered computing could slow its progress somewhat. A comparison to other, even more internationally-monitored technologies, such as nuclear weapons, as well as a review of the experiences of other countries—many of which North Korea learns from either by watching or by sending its experts to study in—suggests the regime could make substantial progress in this area if it commits sufficient time, attention, resources and policy prioritization to it. Indeed, should regime concerns about the loyalty of officials posted overseas or Korean People's Army officers grow in the wake of continued defections, AI solutions could become increasingly appealing for some roles, since as one interviewee we spoke with pointed out, "AI isn't going to defect and doesn't need resources to remain loyal."[164]

Moreover, in the not all that distant future, North Korea may be likely to see AI for offensive cyber operations as of growing value, driven by a combination of factors including advancing technology and know-how, the evolution of adversary cyber defense technologies built around AI, and the costs of developing and maintaining the human capital its cyber forces require. "Offense is still the best return on investment for North Korea," one cybersecurity specialist we spoke with argued, and this could eventually prove to be the case for AI and cyber.[165] Already the regime's leading universities—many of which train its computer hackers and network architects—have demonstrated familiarity with a broad spectrum of AI applications and technologies, suggesting that a substantial portion of the foundation for North Korea to move toward applications in offensive and defensive cyber has already been laid. Pyongyang's social media disinformation campaigns and short message spear phishing ("smishing") techniques in particular may benefit from adoption of natural language processing AI programs that help it tailor messages, including images and voice/sound, to appear more authentic and to dupe foreign targets, especially in light of the limited pool of hackers with high quality foreign language abilities.[166]

Finally, North Korean researchers and hackers are likely to seek ways to exploit vulnerabilities in adversaries' AI-enabled cyber defenses by attempting to steal, corrupt, or reconstruct the data that the systems are built on. "North Korea is a heavy practitioner of denial and deception," one specialist we talked to pointed out, and "AI/ML is very vulnerable to denial and deception," meaning Pyongyang's "first step would almost surely be to try to figure out how [an adversary's] AI system operates."[167] As another expert we spoke with warned, North Korea is exactly the kind of actor that would "deliberately cause the misidentification of a Red Cross truck for a tank...their ability to attack U.S. AI parameters is really a key danger."[168]

While our analysis above has largely focused on AI for cyber, North Korea could certainly seek to apply AI to other aspects of national security, intelligence, military and criminal affairs, and none of the foregoing assessments should be taken to suggest that cyber is the most likely or even necessarily potentially the most impactful area in which Pyongyang might seek to integrate AI. For example, the regime could seek to apply AI to surveillance, sentiment analysis, political loyalty tests, population control, content generation and censorship. Externally, as touched on briefly above, it could move to create more realistic messages in terms of written, spoken, or visual content for spear phishing campaigns and social media disinformation. And in the military realm, sufficiently advanced AI could eventually be applied to fields such as logistics, air defenses, long-range artillery targeting, or terminal guidance for cruise and ballistic missiles.

A number of indicators might give advanced warning that the North was moving to mate AI with its cyber capabilities. First, one might expect to see leadership statements or media reports lauding advances in automated cyber offense or defense, including raising the profiles of individual AI researchers as Kim Jong Un has done for nuclear and missile experts. Separately, academic research by North Korean scholars on AI/ML-based cyber capabilities could signal increased interest by the regime in marrying its cyber efforts with AI. A third possibility would be if the regime appeared to be accelerating its acquisition of training data, advanced computers, or other hardware useful for prosecuting

AI-enabled cyberoperations. Finally, if advances in foreign militaries' or intelligence services' cyber activities signaled an accelerating trend among leading nations in the employment of AI for cyber, this could spur the North to move to keep pace.

In sum, although we do not see North Korea actively engaging in substantial uses of artificial intelligence for offensive or defensive cyber operations, nor do we see it targeting foreign AI systems yet, we do see substantial groundwork upon which the regime could build such an approach should it decide that the incentives it is confronting make such a move advantageous. As such, we conclude that in the future artificial intelligence will likely be seen and employed by Pyongyang as a way to increase the regime's offensive, and potentially also its defensive, cyber capabilities, and may well prove to be a target of its hacking efforts as well, with the result that North Korea may indeed end up using AI for "honing" the regime's "all-purpose sword."

ENDNOTES

[1] For the purposes of this paper, we have chosen to use "artificial intelligence" as a catch-all term that includes the full variety of machine learning, neural networks, deep learning, reinforcement learning, generative adversarial networks, and natural language processing. At various points, below, we use more specific terms to help clarify more delimited applications, including particularly the narrower term "machine learning." Our conception of these terms follows from their use by scholars at the Center for Security and Emerging Technology. See: Center for Security and Emerging Technologies, "Glossary," *n.d.,* accessed September 15, 2021, https://cset.georgetown.edu/glossary/; supplemented by the definitions laid out by IBM, "AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the Difference?" May 27, 2020, https://www.ibm.com/cloud/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks. For shorthand, we often simply use "AI" to refer to the entire field.

[2] Emma Chanlett-Avery, Liana W. Rosen, John W. Rollins, and Catherine A. Theohary, *North Korean Cyber Capabilities: In Brief* (Washington, DC: Congressional Research Service, 2017); Jenny Jun, Scott LaFoy, and Ethan Sohn, *North Korea's Cyber Operations: Strategy and Responses* (Washington, DC: Center for Strategic and International Studies, 2016).

[3] Ben Buchanan, John Bansemer, Dakota Cary, Jack Lucas, and Micah Musser, *Automating Cyber Attacks: Hype and Reality* (Washington, DC: Center for Security and Emerging Technology, November 2020).

[4] Micah Musser and Ashton Garriott, *Machine Learning and Cybersecurity: Hype and Reality* (Washington, DC: Center for Security and Emerging Technology, 2021), https://cset.georgetown.edu/publication/machine-learning-and-cybersecurity/.

[5] Kazumasa Bando, "Is North Korea Next to Unleash AI Hackings After Russia and China?" *Japan Forward,* February 22, 2018, https://japan-forward.com/is-north-korea-next-to-unleash-ai-hackings-after-russia-and-china/.

[6] *Threat Assessment of the U.S. Intelligence Community* (Washington, DC: Office of the Director of National Intelligence, 2021), 15.

[7] *Defense White Paper 2020* (Seoul, Republic of Korea: Ministry of National Defense of the Republic of Korea, 2018), 30.

[8] *Defense of Japan 2020* (Tokyo, Japan: Ministry of Defense of Japan, 2020), 91.

[9] *Defense White Paper 2018* (Seoul, Republic of Korea: Ministry of National Defense of the Republic of Korea, 2018), 27. Another technology outside of cyber that North Korea is investing in is hypersonics. See for example Jeong Tae Joo, "North Korea Forms New Research Center Focused on 'Hypersonic Missiles,'" *Daily NK,* January 6, 2021, https://www.dailynk.com/english/north-korea-forms-new-research-center-focused-on-hypersonic-missiles/.

[10] While we suggest that AI could help make cyber offense and/or cyber defense more effective, to date this has not been demonstrated, and indeed, reliance on AI systems could make cyber offense or defense less effective if the systems are poorly designed. We are describing notional potential futures, but are not predicting these. The authors thank their colleague Chad Heitzenrater for help in identifying and clarifying this point.

[11] Adrian Buzo, *Politics and Leadership in North Korea: The Guerrilla Dynasty, 2nd Ed.* (New York, NY: Routledge, 2017).

[12] *Military and Security Developments Involving the Democratic People's Republic of Korea—Report to Congress* (Washington, DC: Department of Defense, 2017), 5.

[13] For an analysis of North Korea's strategy, see Narushige Michishita, *North Korea's Military-Diplomatic Campaigns, 1966-2008* (New York, NY: Routledge Taylor & Francis, 2010).

[14] Scott W. Harold, "Countering North Korea's Political Warfare," *The Diplomat,* February 10, 2018, https://thediplomat.com/2018/02/countering-north-koreas-political-warfare/.

[15] Jung Pak, *Becoming Kim Jong Un: A Former CIA Officer's Insights into North Korea's Enigmatic Young Dictator* (New York, NY: Ballantine, 2020), 135-136.

[16] Mathew Ha and David Maxwell, *Kim Jong Un's 'All-Purpose Sword': North Korean Cyber-Enabled Economic Warfare* (Washington, DC: Foundation for the Defense of Democracies, 2018).

[17] Anna Fifield, *The Great Successor: The Divines Perfect Destiny of Brilliant Comrade Kim Jong Un* (New York, NY: Public Affairs, 2019), 193-195.

[18] Timothy W. Martin, "North Korea, While Professing Peace, Escalated Cyberattacks on South," *Wall Street Journal,* May 25, 2018, https://www.wsj.com/articles/north-korea-while-professing-peace-escalated-cyberattacks-on-south-1527239057; "N. Korean Hackers Suspected of Continuing Attacks Amid Friendly Inter-Korean Relations," *Yonhap,* July 4, 2018, https://en.yna.co.kr/view/AEN20180705004200320; Mun Dong Hui, "Hackers Use Korea's Divided Family Reunion Plans to Mount 'Spear-Phishing' Attack," *Daily NK,* July 9, 2018, https://www.dailynk.com/english/hackers-use-koreas-divided-family-reunion-plans-to-mount-spear-phishing-attack/; Michael Lee and Park Yong-han, "North Korea's Hackers Target South Korea's Hacks," *JoongAng Ilbo* (English), August 12, 2021, https://koreajoongangdaily.joins.com/2021/08/12/national/northKorea/hacking-North-Korea-media/20210812185107136.html.

[19] RAND Interview #2, cybersecurity expert, May 2021.

[20] RAND Interview #1, North Korea expert, May 2021.

[21] Department of Justice Office of Public Affairs, "Three North Korean Military Hackers Indicted in Wide-Ranging Scheme to Commit Cyberattacks and Financial Crimes Across the Globe," *Department of Justice,* February 17, 2021, https://www.justice.gov/opa/pr/three-north-korean-military-hackers-indicted-wide-ranging-scheme-commit-cyberattacks-and.

[22] For attempts to assess DPRK cyber capabilities, see: Julia Voo, Irfan Hemani, Simon Jones, Winnona DeSombre, Dan Cassidy, and Anna Schwarzenbach, "National Cyber Power Index 2020," The Belfer Center for Science and International Affairs, September 2020, https://www.belfercenter.org/publication/national-cyber-power-index-2020; "Cyber Capabilities and National Power: A Net Assessment," International Institute for Strategic Studies (IISS), June 28, 2021, https://www.iiss.org/blogs/research-paper/2021/06/cyber-capabilities-national-power.

[23] Crowdstrike, 2019 *Global Threat Report: Adversary Tradecraft and the Importance of Speed,* accessed September 11, 2021, https://go.crowdstrike.com/rs/281-OBQ-266/images/Report2019GlobalThreatReport.pdf?lb_email=&utm_source=Marketo&utm_medium=Web&utm_campaign=Threat_Report_2019.

[24] Timothy W. Martin, "North Korea, While Professing Peace, Escalated Cyberattacks on South," *Wall Street Journal,* May 25, 2018, https://www.wsj.com/articles/north-korea-while-professing-peace-escalated-cyberattacks-on-south-1527239057. Such figures should probably be treated with caution as indicators of a general level of commitment rather than regarded as definitive and specific proof of an exact number.

[25] *2020 Defense White Paper* (Seoul, Republic of Korea: ROK Ministry of National Defense, 2021), 29, https://mnd.go.kr/user/mnd/upload/pblictn/PBLICTNEBOOK_202106300300426680.pdf. For many years, estimates of the North's contingent of hackers grew steadily from 600 in 2004 to between 1,000 and 3,000 in 2011 before the figure of 6,800 became the most widely cited following its use in the ROK MND white paper series. Some analysts have speculated that North may even have as many as 30,000 hackers. See "N. Korea's Elite Military Hackers Trained to Attack High Tech Powers," *East-Asia-Intel,* October 12, 2004; "N. Korea's Cyber Warfare Unit in Spotlight after their Attack on S. Korean Bank," *Yonhap,* May 3, 2011, https://en.yna.co.kr/view/AEN20110503010600315; Christine Kim, "Defector Claims North Grooms Hackers," *Joongang Daily,* June 2, 2011, https://koreajoongangdaily.joins.com/2011/06/01/politics/Defector-claims-North-grooms-hackers/2937036.html; "N. Korea Trains Up Hacker Squad," *Chosun Ilbo (English),* March 8, 2011, https://english.chosun.com/site/data/html_dir/2011/03/08/2011030800611.html.

[26] Kaori Yoshida and Yumiko Oshima, "North Korea Sent 1,000 IT Specialists Across the World: UN," *Nikkei Asia,* April 18, 2020, https://asia.nikkei.com/Spotlight/N-Korea-at-crossroads/North-Korea-sent-1-000-IT-specialists-across-the-world-UN-report.

[27] The authors thank Markus Garlauskas for calling their attention to this point.

[28] Ji Young Kong, Kyoung Gon Kim, and Jong In Lim, "The All-Purpose Sword: North Korea," *NATO CCD COE Publications,* May 2019; Headquarters, Department of the Army, "ATP 7-100.2 North Korean Tactics," Army Techniques Publication (Headquarters, Department of the Army, July 24, 2020), https://fas.org/irp/doddir/army/atp7-100-2.pdf.

[29] "North Korean Tactics" (Department of the Army, July 2020), 277, https://www.documentcloud.org/documents/7038686-US-Army-report-on-North-Korean-military.html.

[30] Christine Kim, "Defector Claims North Grooms Hackers," *JoongAng Ilbo (English edition),* June 1, 2011, https://koreajoongangdaily.joins.com/news/article/Article.aspx?aid=2937036.

[31] Ha and Maxwell, "Kim Jong Un's All-Purpose Sword."

[32] Adapted from Kim Chong Woo and Carolina Polito, "The Evolution of North Korean Cyber Threats," Asan Institute for Policy Studies, February 20, 2019, https://en.asaninst.org/contents/the-evolution-of-north-korean-cyber-threats/.

[33] Here we adopt the STRIDE framework for evaluating cyber threats put forward by Microsoft. See "The STRIDE Threat Model," *Microsoft,* November 12, 2009, https://docs.microsoft.com/en-us/previous-versions/commerce-server/ee823878(v=cs.20)?redirectedfrom=MSDN.

[34] "Ten Days of Rain," *McAfee,* 2011, https://www.mcafee.com/wp-content/uploads/2011/07/McAfee-Labs-10-Days-of-Rain-July-2011.pdf.

[35] David Martin, "Tracing the Lineage of DarkSeoul," SANS, March 4, 2016, https://www.sans.org/white-papers/36787/.

[36] Dmitry Tarakanov, "The 'Kimsuky' Operation: A North Korean APT?" *SecureList by Kaspersky,* September 11, 2013, https://securelist.com/the-kimsuky-operation-a-north-korean-apt/57915/.

[37] Jeyup S. Kwaak, "North Korea Blamed for Nuclear Power Plant Hack," *Wall Street Journal,* March 17, 2015, https://www.wsj.com/articles/north-korea-blamed-for-nuclear-power-plant-hack-1426589324.

[38] Michael Balsamo, "North Korean Programmer Charged in Sony Hack, WannaCry Attack," *PBS News Hour,* September 6, 2018, https://www.pbs.org/newshour/nation/north-korean-programmer-charged-in-sony-hack-wannacry-attack.

[39] Jonathan Shieber, "North Korean Hackers Stole South Korean and U.S. War Plans," *Tech Crunch,* October 10, 2017, https://techcrunch.com/2017/10/10/report-north-korean-hackers-stole-south-korean-and-u-s-war-plans/.

[40] "FASTCash 2.0: North Korea's BeagleBoyz Robbing Banks," Cybersecurity and Infrastructure Security Agency, August 26, 2020, https://us-cert.cisa.gov/ncas/alerts/aa20-239a.

[41] Timothy Martin, Eun-Young Jeong, and Steven Russolillo, "North Korea Is Suspected in Bitcoin Heist," *Wall Street Journal,* December 20, 2017, https://www.wsj.com/articles/north-korea-is-suspected-in-bitcoin-heist-1522303177.

[42] Kwaak, "North Korean Programmer Charged."

[43] Jon Russell, "Korean Crypto Exchange Coinrail Loses Over $40M in Tokens Following a Hack," *TechCrunch,* June 10, 2018, https://techcrunch.com/2018/06/10/korean-crypto-exchange-coinrail-loses-over-40m-in-tokens-following-a-hack/; Jon Russell, "Korean Crypto Exchange Bithumb Says It Lost Over $30M Following a Hack," *TechCrunch,* June 19, 2018, https://techcrunch.com/2018/06/19/korean-crypto-exchange-bithumb-says-it-lost-over-30m-following-a-hack/.

[44] Ryan Sherstobitoff, "Analyzing Operation GhostSecret: Attack Seeks to Steal Data Worldwide," *McAfee,* April 24, 2018, accessed November 23, 2021 at: https://www.mcafee.com/blogs/other-blogs/mcafee-labs/analyzing-operation-ghostsecret-attack-seeks-to-steal-data-worldwide/.

[45] Chong Woo Kim and Carolina Polito, "The Evolution of North Korean Cyber Threats," Issue Briefs (Asan Institute for Policy Studies, February 20, 2019), http://en.asaninst.org/contents/the-evolution-of-north-korean-cyber-threats/. The primary focus of the regime's cyber activities appears to be on fundraising, though at times other priorities assume greater importance. Additionally, there can be tensions between the regime's various goals for using cyber, for example between raising revenues or conducting military or intelligence operations. And while the regime's cyber activities may aim at driving a wedge between the ROK and the U.S., it is possible that the threat instead leads to greater allied cooperation. The authors thank an anonymous reviewer for suggesting that the author clarify these points.

[46] Ju-min Park and Meeyoung Cho, "South Korea Blames North Korea for December Hack on Nuclear Operator," *Reuters,* March 17, 2015, https://www.reuters.com/article/us-nuclear-southkorea-northkorea/south-korea-blames-north-korea-for-december-hack-on-nuclear-operator-idUSKBN0MD0GR20150317.

[47] "Alert (AA20-106A): Guidance on the North Korean Cyber Threat," U.S. Cybersecurity and Infrastructure Security Agency, June 23, 2020, https://us-cert.cisa.gov/ncas/alerts/aa20-106a; "Alert (AA20-239A): FASTCash 2.0: North Korea's BeagleBoyz Robbing Banks," U.S. Cybersecurity and Infrastructure Security Agency, October 24, 2020, https://us-cert.cisa.gov/ncas/alerts/aa20-239a;"Alert (TA18-275A): HIDDEN COBRA – FASTCash Campaign," U.S. Cybersecurity and Infrastructure Security Agency, December 21, 2018, https://us-cert.cisa.gov/ncas/alerts/TA18-275A; Choe Sang-hun, "North Korea Stole Data of Millions of Online Consumers, South Says," *New York Times,* July 28, 2016, https://www.nytimes.com/2016/07/29/world/asia/north-korea-hacking-interpark.html.

[48] Tara Seals, "Banco de Chile Wiper Attack Just a Cover for $10M SWIFT Heist," *Threat Post,* June 13, 2018, https://threatpost.com/banco-de-chile-wiper-attack-just-a-cover-for-10m-swift-heist/132796/.

[49] For more information on the WannaCry ransomware, see: "What You Need to Know about the WannaCry Ransomware," *Symantec,* October 23, 2017, https://symantec-enterprise-blogs.security.com/blogs/threat-intelligence/wannacry-ransomware-attack. For more on North Korea's record of using worms, refer to: Chris Doman, "North Korean Cyber-Attacks and Collateral Damage," AT&T, February 15, 2018, https://cybersecurity.att.com/blogs/security-essentials/north-korean-cyber-attacks-and-collateral-damage.

[50] Catalin Cimpanu, "TrickBot gang is now a malware supplier for North Korean hackers," *ZDNet,* December 11, 2019, https://www.zdnet.com/article/trickbot-gang-is-now-a-malware-supplier-for-north-korean-hackers/; "Alert (AA20-106A): Guidance on the North Korean Cyber Threat," U.S. Cybersecurity and Infrastructure Security Agency, June 23, 2020, https://us-cert.cisa.gov/ncas/alerts/aa20-106a; "North Korean hackers are skimming US and European shoppers," Sansec, July 6, 2020, https://sansec.io/research/north-korea-magecart.

[51] Ivan Kwiatkowski, Pierre Delcher, and Felix Aime, "Lazarus on the hunt for big game," *Kaspersky Secure List,* July 28, 2020, https://securelist.com/lazarus-on-the-hunt-for-big-game/97757/.

[52] "A Brief Look At North Korean Cryptography," *Kryptos Logic,* July 3, 2018, https://www.kryptoslogic.com/blog/2018/07/a-brief-look-at-north-korean-cryptography/.

[53] Martyn Williams, "Catch Me If You Can: North Korea Works to Improve Communications Security," *38North,* April 12, 2017, https://www.38north.org/2017/04/mwilliams041217/, Martyn Williams, "North Korea's Koryolink: Built for Surveillance and Control," 38North, July 22, 2019, https://www.38north.org/2019/07/mwilliams072219/.

[54] Eul-Chul Lim, "North Korea's Fourth Industrial Revolution: Strategy, Direction, Method, and Progress (북한의 4차 산업혁명 : 대응전략, 추진방식과 성과)," Donga Research Institute, 2019, https://www.dbpia.co.kr/Journal/articleDetail?nodeId=NODE09251604.

[55] Pratik Jakhar, "North Korea's High-Tech Pursuits: Propaganda or Progress?" *BBC News,* December 15, 2018, https://www.bbc.com/news/world-asia-46563454.

[56] Jingyu Kang, "North Korea 'In Age of Artificial Intelligence, Data is more Valuable than Gold and Gas'북한 인공지능 시대 데이터가 금, 원유보다 중요,', November 1, 2019, http://www.nkeconomy.com/news/articleView.html?idxno=2158.

[57] Elizabeth Shim, "North Korea University to Teach Artificial Intelligence, State Media Says," *United Press International,* June 23, 2019, https://www.upi.com/Top_News/World-News/2019/06/23/North-Korea-university-to-teach-artificial-intelligence-state-media-says/4261561308306/.

[58] Jinkyu Kang, "North Korea, '[We] Should Provide the Latest Scientific Data to Scientists and Researchers'," ("북한 '과학자, 기술자들에게 최신 과학기술자료 제공해야,'") *NK Economy*, August 5, 2021, https://www.nkeconomy.com/news/articleView.html?idxno=4519.

[59] Kang, "North Korea, '[We] Should Provide the Latest Scientific Data to Scientists and Researchers.'"

[60] Choe Sang-Hun, Motoko Rich, Natalie Reneau, and Audrey Carlsen, "Rocket Men: The Team Building North Korea's Nuclear Missile," *New York Times*, December 15, 2017, https://www.nytimes.com/interactive/2017/12/15/world/asia/north-korea-scientists-weapons.html; Sung-hui Moon, "More North Koreans Become Scientists to Reap Privileges From The Regime," *Radio Free Asia*, May 18, 2017, https://www.rfa.org/english/news/korea/more-north-koreans-become-scientists-to-reap-privileges-from-the-regime-05182017162458.html.

[61] RAND interviews with North Korea experts (Interview #4A and #4B) and cybersecurity expert (Interview #2), May 2021.

[62] "S/2021/211," *Final Report of the Panel of Experts* (New York, NY: United Nations Security Council, March 4, 2021), https://undocs.org/S/2021/211.

[63] Detailed list of *Kim Il Sung University Gazette (hakbo)* articles consulted for this study is available upon request; journal articles from PUST and other North Korean universities were inaccessible.

[64] Pak Song Ho (박성호) and Hwang Chol Jin (황철진), "Performance Improvement by Attribute Selection in the Network Intrusion Detection System" ("망침입검출에서 속성선택에 의한 성능개선"), *Kim Il Sung University Gazette* 64, no. 2, 2018.

[65] Sim Yun Go (심윤거) and Pak Myong Suk (박명숙), "A Fuzzy Reasoning Method and IKEv2 Protocol Design for Preventing IP Spoofing DoS Attack," ("IP속임DoS공격방지를 위한 모호추론방법과 IKEv2규약설계") *Kim Il Sung University Gazette* 65, no. 3, 2019.

[66] Jinkyu Kang, "Is North Korea Increasing Exports of AI-based Biometric Recognition Systems?" ("북한, AI기반 생체인증 제품 수출 강화하나") *NK Economy*, April 4, 2021, http://www.nkeconomy.com/news/articleView.html?idxno=4147.

[67] SongIl Choe, Bo Li, IlNam Ri, ChangSu Paek, JuSong Rim, and SuBom Yun, "Improved Hybrid Symbiotic Organism Search Task-Scheduling Algorithm for Cloud Computing," *KSII Transactions on Internet and Information Systems* 12, no. 8 (August 31, 2018), https://doi.org/10.3837/tiis.2018.08.001.

[68] Jinkyu Kang, "North Korea Working on Clouding Computing Infrastructure" ("북한 정보과학기술연구소, 클라우드 컴퓨팅 시스템 구축") *NK Economy*, September 19, 2019, https://www.nkeconomy.com/news/articleView.html?idxno=2081.

[69] Ri Chung Il (리충일), Kim Sun Dol (김순돌), and Choe Chol (최철), "A Method of Real-time Vehicle License Plate Location Using Deep Neural Network," ("심층신경망을 리용한 실시간차번호판령역검출의 한가지 방법") *Kim Il Sung University Hakbo* 66, no. 4, 2020.

[70] Jinkyu Kang, "North Korea Uses MIT Data to Conduct AI Research on X-ray Technology," ("북한, 미국 MIT 데이터로 X레이 AI 분석 기술 개발,") *NK Economy*, August 15, 2021, https://www.nkeconomy.com/news/articleView.html?idxno=4540.

[71] Alistair E. W. Johnson, Tom Pollard, Roger Mark, Seth Berkowitz, and Steven Horng, "The MIMIC-CXR Database," Physionet.org, 2019, https://doi.org/10.13026/C2JT1Q.

[72] Ri, Kim, and Choe, "A Method of Real-time Vehicle License Plate Location Using Deep Neural Network"; Jinkyu Kang, "North Korea Uses GoogLeNet for Image Processing," ("북한, 인공지능 구글넷 기술로 영상 분석한다") *NK Economy*, May 25, 2020, https://www.nkeconomy.com/news/articleView.html?idxno=3100.

[73] See, for example: John A. Mathews and Dong-Sung Cho, *Tiger Technology: The Creation of a Semiconductor Industry in East Asia* (Cambridge, UK: Cambridge University Press, 2007); Dan Ciuriak, "How Can Companies in Emerging Markets Acquire New Technology?" Davos World Economic Forum, April 10, 2015, https://www.weforum.org/agenda/2015/04/how-can-companies-in-emerging-markets-acquire-new-technology/; Carl Dahlman, "Technology, Globalization, and International Competitiveness: Challenges for Developing Countries," in *Industrial Development for the 21st Century: Sustainable Development Perspectives* (New York, NY: United Nations Department of Economic and Social Affairs, 2007), https://www.un.org/esa/sustdev/publications/industrial_development/1_2.pdf.

[74] RAND Interview #2, cybersecurity expert, May 2021.

[75] Jonathan Pollack, *No Exit: North Korea, Nuclear Weapons and International Security* (London, UK: International Institute for Strategic Studies, 2011).

[76] Priscilla Moriuchi and Fred Wolens, "North Korea Relies on American Technology for Internet Operations," Insikt Group, June 6, 2018, https://www.recordedfuture.com/north-korea-internet-operations/.

[77] John Park and Jim Walsh, *Stopping North Korea, Inc.: Sanctions Effectiveness and Unintended Consequences* (Cambridge, MA: MIT Security Studies Program, August 2016).

[78] Jan Ransom, "He Gave a Cryptocurrency Talk in North Korea. Then the U.S. Arrested Him," *New York Times*, December 2, 2019, https://www.nytimes.com/2019/12/02/nyregion/north-korea-virgil-griffin-cyptocurrency-arrest.html.

[79] "United States Citizen Pleads Guilty to Conspiring to Assist North Korea in Evading Sanctions," U.S. Department of Justice, September 27, 2021, https://www.justice.gov/usao-sdny/pr/united-states-citizen-pleads-guilty-conspiring-assist-north-korea-evading-sanctions.

[80] RAND Interview #2, cybersecurity expert, May 2021.

[81] RAND Interview #3, North Korea expert, April 2021.

[82] Kang Seung-woo, "'North Korea Was Once AI Powerhouse'," *The Korea Times,* October 17, 2017, http://www.koreatimes.co.kr/www/tech/2018/05/133_237814.html.

[83] One of the possible targets of North Korean theft could be large-scale databases that would be used for training AI/ML algorithms. As noted below, some training data sets are publicly available, but others, where privately held, may be of greater value to the regime and might be targeted for theft. The authors thank an anonymous reviewer for suggesting that they make this point more explicit.

[84] Han Sang Mi, "North Korea sends 50-60 STEM Talents Abroad for Cyber Agent Training," Voice of America, June 14, 2016, https://www.voakorea.com/korea/korea-politics/3375411. It is unclear from open-source reporting exactly what type of cyber capabilities these students are being trained in or studying.

[85] Andrew Salmon, "North Korea and the 'Dead Hand' Nuclear Strategy," *Asia Times,* February 25, 2020, https://asiatimes.com/2020/02/north-korea-and-the-dead-hand-nuclear-strategy/; Jeremy Page and Alastair Gale, "Behind North Korea's Nuclear Advance: Scientists Who Bring Technology Home," *Wall Street Journal,* September 6, 2017, https://www.wsj.com/articles/behind-north-koreas-nuclear-advance-scientists-who-bring-technology-home-1504711605.

[86] Jeremy Page and Alastair Gale, "Behind North Korea's Nuclear Advance: Scientists Who Bring Technology Home," Wall Street Journal, September 6, 2017, https://www.wsj.com/articles/behind-north-koreas-nuclear-advance-scientists-who-bring-technology-home-1504711605.

[87] Kim's research focused on the application of deep learning to use neuroscience to improve commercial marketing to consumers. Kim Chungson [金忠星], "Research on the Consumer Preference Prediction System in Neuromarketing" ["神经营销的消费者偏好预测系统研究"], PhD dissertation, Harbin Institute of Technology [哈尔滨工业大学], 2019; Kim Chungson [金忠星] and Li Dong [李东], "Deep learning neural network model for consumer preference prediction" ["消费者偏好预测的深度学习神经网络模型"], *Journal of Computer Applications* [计算机应用] 39:7, (2019):1888-1893.

[88] For one example list, see: "Top 10 Datasets for Cybersecurity Projects," *Analytics India Magazine,* October 28, 2020, https://analyticsindiamag.com/top-10-datasets-for-cybersecurity-projects/.

[89] *Final report of the Panel of Experts submitted pursuant to resolution 2464 (2019),* report number S/2020/151, (New York, NY: United Nations, March 2, 2020), 50, https://undocs.org/S/2020/151.

[90] Ransom, "He Gave a Cryptocurrency Talk in North Korea. Then the U.S. Arrested Him."

[91] Jeremy Page, "North Korea Is Making Millions of Dollars Selling Power to China," *Wall Street Journal,* March 16, 2018, https://www.wsj.com/articles/north-korea-is-making-millions-of-dollars-selling-power-to-china-1521192603. *U.S. Energy Information Agency* estimates DPRK total power production at 11 mwh in 2015: "North Korea," U.S. Energy Information Agency, last updated June 2018, https://www.eia.gov/international/analysis/country/PRK.

[92] Marin Vlastelica Pogančić, "The Carbon Footprint of AI Research," Towards Data Science, October 1, 2019, https://towardsdatascience.com/the-carbon-footprint-of-ai-research-812d9c974a5c; Rohollah Faghihi, "Iran: Is Chinese Bitcoin mining behind country's massive power blackouts?" *Middle East Eye,* January 23, 2021, https://www.middleeasteye.net/news/iran-china-bitcoin-massive-power-blackouts.

[93] J Nicholas Hoover, "NSA Building $896.5 Million Supercomputing Center," *Information Week,* April 21, 2011, https://www.informationweek.com/architecture/nsa-building-$8965-million-supercomputing-center/d/d-id/1097313; Karen Hao, "Training a single AI model can emit as much carbon as five cars in their lifetimes," *MIT Technology Review,* June 6, 2019, https://www.technologyreview.com/2019/06/06/239031/training-a-single-ai-model-can-emit-as-much-carbon-as-five-cars-in-their-lifetimes/.

[94] *How North Korea Revolutionized the Internet as a Tool for Rogue Regimes,* Insikt Group, February 9, 2020, https://www.recordedfuture.com/north-korea-internet-tool/.

[95] Sam Kim, "Inside North Korea's Hacker Army," *Bloomberg,* February 7, 2018, https://www.bloomberg.com/news/features/2018-02-07/inside-kim-jong-un-s-hacker-army. ATP 7-100.2: North Korean Tactics, Headquarters, Department of the Army, July 2020, http://www.documentcloud.org/documents/7038686-US-Army-report-on-North-Korean-military.html.

[96] The reports add that despite the regime's guidance, some of these overseas IT workers do moonlight as hackers anyway. See: *Final report of the Panel of Experts submitted pursuant to resolution* 2464 (2019), report number S/2020/151, (New York, NY: United Nations, March 2, 2020).

[97] "N. Korea ranks near bottom among countries in Internet speed," *Korea Times,* June 8, 2016, http://www.koreatimes.co.kr/www/nation/2020/03/103_206499.html.

[98] Tai Wei Lim, "North Korea's Artificial Intelligence (A.I.)," *North Korean Review,* 15, no. 2 (Fall 2019): 97-103.

[99] Jason Murdock, "North Korea: Kim Jong-un hates the United States but loves his MacBook," *International Business Times,* February 12, 2016, https://www.ibtimes.co.uk/north-korea-kim-jong-un-hates-united-states-loves-his-macbook-1543508.

[100] Jeff Heaton, "Building a $5,000 Machine Learning Workstation with an NVIDIA TITAN RTX and RYZEN ThreadRipper," *Towards Data Science,* July 22, 2020, https://towardsdatascience.com/building-a-5-000-machine-learning-workstation-with-an-nvidia-titan-rtx-and-ryzen-threadripper-46c49383fdac.

[101] RAND Interview #2, cybersecurity expert, May 2021.

[102] See Mun Dong Hui, "N. Korea Releases Details about Development of 'Real-Time License Plate Recognition Technology,'" *Daily NK,* April 21, 2021, https://www.dailynk.com/english/north-korea-releases-details-about-development-real-time-license-plate-recognition-technology/. The computer science labs at Kim Il Sung University use GTX-980 GPUs manufactured by NVIDIA in 2014—this is a relatively older model compared to its newest releases such as the RTX 3000-series. Deep learning processes are difficult to conduct with a GTX-980.

[103] RAND Interviews #1 and #4A, both North Korea experts, May 2021.

[104] RAND Interview #1, North Korea expert, May 2021.

[105] "Google Says North Korean State Hackers are Targeting Security Researchers on Social Media," CNBC, January 26, 2021, https://www.cnbc.com/2021/01/26/north-korean-hackers-targeting-security-researchers-on-twitter.html.

[106] RAND Interviews #1 and #4A, both North Korea experts, May 2021.

[107] Gregory C. Allen, *Understanding China's AI Strategy* (Washington, DC: Center for a New American Security, 2019); Elsa B. Kania, *Chinese Military Innovation in Artificial Intelligence: Hearing of the U.S.-China Economic and Security Review Commission* (Washington, DC: Center for a New American Security, 2019); Ryan Fedasiuk, *Chinese Perspectives on AI and Future Military Capabilities* (Washington, DC: Center for Strategic and Emerging Technologies, August 2020); Kai-Fu Lee, *AI Superpowers: China, Silicon Valley, and the New World Order* (New York, NY: Houghton Mifflin Harcourt, 2018).

[108] "The Global AI Talent Tracker," *Macro Polo,* n.d., accessed September 9, 2021 at: https://macropolo.org/digital-projects/the-global-ai-talent-tracker/.

[109] Ashwin Acharya and Zachary Arnold, *Chinese Public AI R&D Spending: Provisional Findings* (Washington, DC: Center for Security and Emerging Technology, December 2019), https://cset.georgetown.edu/research/chinese-public-ai-rd-spending-provisional-findings/.

[110] Katie Benner, "North Koreans Accused of Laundering $2.5 Billion for Nuclear Program," *New York Times,* May 28, 2020, https://www.nytimes.com/2020/05/28/us/politics/north-korea-money-laundering-nuclear-weapons.html.

[111] Scott W. Harold, Nathan Beauchamp-Mustafaga, and Jeffrey W. Hornung, *Chinese Disinformation Efforts on Social Media* (Santa Monica, CA: The RAND Corporation, 2021); Jessica Brandt and Torrey Taussig, "The Kremlin's Disinformation Playbook Goes to Beijing," Brookings Institution, May 19, 2020, https://www.brookings.edu/blog/order-from-chaos/2020/05/19/the-kremlins-disinformation-playbook-goes-to-beijing/.

[112] Robert King, "Fake News from Pyongyang! How North Korea Is Using the Internet," *The Peninsula,* Korean Economic Institute of America (blog), May 27, 2020, https://keia.org/the-peninsula/fake-news-from-pyongyang-how-north-korea-is-using-the-internet/; Mathew Ha, "North Korea Turns to Cyber Disinformation Attacks Amid Global Coronavirus Outbreak," Foundation for the Defense of Democracies, April 1, 2020, https://www.fdd.org/analysis/2020/04/01/north-korea-turns-to-cyber-disinformation-attacks-amid-global-coronavirus-outbreak/; Jason Bartlett, "From Ethnic Nationalism to Social Media: How North Korea Leverages Its Soft Power Abroad," *The Diplomat,* August 17, 2021, https://thediplomat.com/2021/08/from-ethnic-nationalism-to-social-media-how-north-korea-leverages-its-soft-power-abroad/.

[113] "Putin: Leader in artificial intelligence will rule world," *Associated Press,* September 4, 2017, https://www.cnbc.com/2017/09/04/putin-leader-in-artificial-intelligence-will-rule-world.html.

[114] Andy Greenberg, "The Untold Story of NotPetya, the Most Devastating Cyberattack," *Wired,* August 22, 2018, https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/.

[115] Joe Cheravitch and Bilyana Lilly, "Russia's Cyber Limitations in Personnel and Innovation, Their Potential Impact on Future Operations, and How NATO and Its Members Can Respond," in A. Ertan, K. Floyd, P. Pernik, T. Stevens, eds., *Cyber Threats and NATO 2030: Horizon Scanning and Analysis,* NATO Cooperative Cyber Defense Centre of Excellence, 2020, https://ccdcoe.org/uploads/2020/12/Cyber-Threats-and-NATO-2030_Horizon-Scanning-and-Analysis.pdf.

[116] Nikolai Markotkin and Elena Chenenko, "Developing Artificial Intelligence in Russia: Objectives and Reality," Carnegie Endowment for International Peace Moscow Center, August 5, 2020, https://carnegiemoscow.org/commentary/82422.

[117] Alina Polyakova, "Weapons of the weak: Russia and AI-driven asymmetric warfare," Brookings Institution, November 15, 2018, https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/, https://www.brookings.edu/research/weapons-of-the-weak-russia-and-ai-driven-asymmetric-warfare/; Mathew Ha, "North Korea Turns to Cyber Disinformation Attacks Amid Global Coronavirus Outbreak," Foundation for the Defense of Democracies, April 1, 2020, https://www.fdd.org/analysis/2020/04/01/north-korea-turns-to-cyber-disinformation-attacks-amid-global-coronavirus-outbreak/; Robert King, "Fake News from Pyongyang! How North Korea is Using the Internet," *The Peninsula,* Korean Economic Institute of America (Blog), May 20, 2020, https://keia.org/the-peninsula/fake-news-from-pyongyang-how-north-korea-is-using-the-internet/.

[118] Mahmoud Pargoo, "Sanctions Propel Iran in the Global Race for Terminator-like AI," Atlantic Council, April 2, 2019, https://www.atlanticcouncil.org/blogs/iransource/sanctions-propel-iran-in-the-global-race-for-terminator-like-ai/.

[119] Nicole Perlroth and Clifford Krauss, "A Cyberattack in Saudi Arabia Had a Deadly Goal. Experts Fear Another Try," *New York Times,* March 15, 2018, https://www.nytimes.com/2018/03/15/technology/saudi-arabia-hacks-cyberattacks.html; Andy Greenberg, "A Notorious Iranian Hacking Crew Is Targeting Industrial Control Systems," Wired, November 20, 2019,: https://www.wired.com/story/iran-apt33-industrial-control-systems/.

[120] Patrick Thibodeau, "Iran Says It's Building an AI Supercomputer, Despite Sanctions," *Tech Target,* August 23, 2019, https://searchdatacenter.techtarget.com/news/252469119/Iran-says-its-building-an-AI-supercomputer-despite-sanctions.

[121] Zachary Fryer-Biggs, "Twilight of the Human Hacker: Secretive Pentagon Research Program Seeks to Replace Hackers with AI," *Public Integrity,* September 13, 2020, https://publicintegrity.org/national-security/twilight-of-the-human-hacker-cyberwarfare/.

[122] Cho Jin-young, "S. Korea, U.S. Work Together to Develop AI-Based Cyber Security Technology," *Business Korea,* May 3, 2016, http://www.businesskorea.co.kr/news/articleView.html?idxno=14572.

[123] "Japan Embraces AI Tools to Fight Cyberattacks with US$237 Million Investment," *CISOMAG,* April 6, 2020, https://cisomag.eccouncil.org/japan-embraces-ai-tools-to-fight-cyberattacks-with-us237-mn-investment/.

[124] The authors thank their colleague Chad Heitzenrater for pointing this out.

[125] RAND Interview #2, cybersecurity expert, May 2021.

[126] FM 3-28 defines Offensive Cyberspace Operations as "cyberspace operations intended to project power by the application of force in or through cyberspace," Department of the Army, "FM 3-38: Cyber Electromagnetic Activities," February 12, 2014, https://irp.fas.org/doddir/army/fm3-38.pdf. At various points below we also refer to influence operations such as online disinformation and propaganda campaigns that do not specifically rely on network exploitation.

[127] Researchers at Mandiant find North Korea has executed three zero-day attacks in the past four years. See: Kathleen Metrick, Parnian Najafi, and Jared Semrau, "Zero Day Exploitation Increasingly Demonstrates Access to Money, Rather than Skill—Intelligence for Vulnerability Management, Part One," *Mandiant,* April 6, 2020, https://www.mandiant.com/resources/zero-day-exploitation-demonstrates-access-to-money-not-skill.

[128] "What Is a Cyberattack? - Most Common Types," Cisco, accessed September 15, 2021, https://www.cisco.com/c/en/us/products/security/common-cyberattacks.html; "What Is Spear Phishing? - Definition," Kaspersky, January 13, 2021, https://usa.kaspersky.com/resource-center/definitions/spear-phishing; "What Is an Advanced Persistent Threat (APT)?" Kaspersky, June 11, 2021, https://www.kaspersky.com/resource-center/definitions/advanced-persistent-threats; "What Is Ransomware?," McAfee, accessed September 15, 2021, https://www.mcafee.com/enterprise/en-us/security-awareness/ransomware.html.

[129] Catalin Cimpanu, "TrickBot Gang Is Now a Malware Supplier for North Korean Hackers," *ZDNet,* December 11, 2019, https://www.zdnet.com/article/trickbot-gang-is-now-a-malware-supplier-for-north-korean-hackers/.

[130] RAND Interview #5, cybersecurity and AI specialist, April 2021.

[131] Such circumstances are by no means unique and are often common when new defenses are deployed without their in-the-wild capabilities having been fully understood and tested.

[132] Micah Musser and Ashton Garriott, *Machine Learning and Cybersecurity: Hype and Reality.*

[133] "How Hackers Are Using AI Technologies to Develop Intelligent Malware," *CISOMAG,* December 9, 2019, https://cisomag.eccouncil.org/hackers-using-ai/.

[134] Robert Fay and Wallace Trenholm, "The Cyber Security Battlefield: AI Technology Offers Both Opportunities and Threats," in *Governing Cyberspace during a Crisis in Trust* (Waterloo, Canada: Centre for International Governance Innovation, 2019), https://www.jstor.org/stable/resrep26129.11.

[135] Anirban Chakraborty, Manaar Alam, Vishal Dey, Anupam Chattopadhyay, and Debdeep Mukhopadhyay, "Adversarial Attacks and Defences: A Survey," *ArXiv:1810.00069 [Cs, Stat],* September 28, 2018, http://arxiv.org/abs/1810.00069; Victor Shepardson, Gary McGraw, Harold Figueroa, and Richie Bonett, "A Taxonomy of ML Attacks," *BerryvilleIML,* May 2019, https://berryvilleiml.com/taxonomy/.

[136] For more information about adversarial machine learning, refer to: Xiaoyong Yuan, Pan He, Qile Zhu, and Xiaolin Li, "Adversarial Examples: Attacks and Defenses for Deep Learning," *ArXiv:1712.07107 [Cs, Stat],* July 6, 2018, http://arxiv.org/abs/1712.07107; Chakraborty, Alam, Dey, Chattopadhyay, and Mukhopadhyay, "Adversarial Attacks and Defences: A Survey"; Nicolas Papernot, Patrick McDaniel, Ian Goodfellow, Somesh Jha, Z. Berkay Celik, and Ananthram Swami, "Practical Black-Box Attacks against Machine Learning," in *Proceedings of the 2017 ACM on Asia Conference on Computer and Communications Security,* (ASIA CCS '17: ACM Asia Conference on Computer and Communications Security, Abu Dhabi United Arab Emirates: ACM, 2017), 506-19, https://doi.org/10.1145/3052973.3053009.

[137] Alex Polyakov, "How to Attack Machine Learning (Evasion, Poisoning, Inference, Trojans, Backdoors)," *TowardsDataScience.com,* August 6, 2019, https://towardsdatascience.com/how-to-attack-machine-learning-evasion-poisoning-inference-trojans-backdoors-a7cb5832595c.

[138] Bojan Kolosnjaji, Ambra Demontis, Battista Biggio, Davide Maiorca, Giorgio Giacinto, Claudia Eckert, and Fabio Roli, "Adversarial Malware Binaries: Evading Deep Learning for Malware Detection in Executables," in *2018 26th European Signal Processing Conference (EUSIPCO),* (2018 26th European Signal Processing Conference (EUSIPCO), Rome: IEEE, 2018), 533-37, https://doi.org/10.23919/EUSIPCO.2018.8553214; Felix Kreuk, Assi Barak, Shir Aviv-Reuven, Moran Baruch, Benny Pinkas, and Joseph Keshet, "Deceiving End-to-End Deep Learning Malware Detectors Using Adversarial Examples," *ArXiv:1802.04528 [Cs],* January 10, 2019, http://arxiv.org/abs/1802.04528; Octavian Suciu, Scott E. Coull, and Jeffrey Johns, "Exploring Adversarial Examples in Malware Detection," *ArXiv:1810.08280 [Cs, Stat],* April 13, 2019, http://arxiv.org/abs/1810.08280.

[139] "Cylance, I Kill You!" *Skylight* (blog), September 7, 2019, https://skylightcyber.com/2019/07/18/cylance-i-kill-you/.

[140] "Cylance, I Kill You!"

[141] Alexey Antonov and Alexey Kogtenkov, "How to Confuse Antimalware Neural Networks. Adversarial Attacks and Protection," Securelist (blog), June 23, 2021, https://securelist.com/how-to-confuse-antimalware-neural-networks-adversarial-attacks-and-protection/102949/.

[142] Bhargav Kuchipudi, Ravi Teja Nannapaneni, and Qi Liao, "Adversarial Machine Learning for Spam Filters," in *Proceedings of the 15th International Conference on Availability,* Reliability and Security (ARES 2020: The 15th International Conference on Availability, Reliability and Security, Virtual Event Ireland: ACM, 2020), 1-6, https://doi.org/10.1145/3407023.3407079; Chenran Wang, Danyi Zhang, Suye Huang, Xiangyang Li, and Leah Ding, "Crafting Adversarial Email Content against Machine Learning Based Spam Email Detection," in *Proceedings of the 2021 International Symposium on Advanced Security on Software and Systems* (ASIA CCS '21: ACM Asia Conference on Computer and Communications Security, Virtual Event Hong Kong: ACM, 2021), 23-28, https://doi.org/10.1145/3457340.3458302.

[143] Dmitri Alperovitch and Ian Ward, "How Should the U.S. Respond to the SolarWinds and Microsoft Exchange Hacks *Lawfare,* March 12, 2021, https://www.lawfareblog.com/how-should-us-respond-solarwinds-and-microsoft-exchange-hacks.

[144] Jack Kim, "North Korea Mounts Long-Running Hack of South Korea Computers, Says Seoul," *Reuters,* June 13, 2016, https://www.reuters.com/article/us-northkorea-southkorea-cyber/north-korea-mounts-long-running-hack-of-south-korea-computers-says-seoul-idUSKCN0YZ0BE.

[145] Charlie Osborne, "Updated Kaseya Ransomware Attack FAQ: What We Know Now," *ZDNet,* July 23, 2021, https://www.zdnet.com/article/updated-kaseya-ransomware-attack-faq-what-we-know-now/.

[146] Nicholas Weaver, "The Microsoft Exchange Hack and the Great Email Robbery," *Lawfare* (blog), March 9, 2021, https://www.lawfareblog.com/microsoft-exchange-hack-and-great-email-robbery.

[147] Hongsheng Hu, Zoran Salcic, Gillian Dobbie, and Xuyun Zhang, "Membership Inference Attacks on Machine Learning: A Survey," *ArXiv:2103.07853 [Cs],* March 16, 2021, http://arxiv.org/abs/2103.07853; Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov, "Membership Inference Attacks against Machine Learning Models," *ArXiv:1610.05820 [Cs, Stat],* March 31, 2017, http://arxiv.org/abs/1610.05820; Liwei Song, Reza Shokri, and Prateek Mittal, "Membership Inference Attacks Against Adversarially Robust Deep Learning Models," *2019 IEEE Security and Privacy Workshops (SPW)* (San Francisco, CA: IEEE, 2019), 50-56, https://doi.org/10.1109/SPW.2019.00021.

[148] Giuseppe Ateniese, Giovanni Felici, Luigi V. Mancini, Angelo Spognardi, Antonio Villani, and Domenico Vitali, "Hacking Smart Machines with Smarter Ones: How to Extract Meaningful Data from Machine Learning Classifiers," *ArXiv:1306.4447 [Cs, Stat],* June 19, 2013, http://arxiv.org/abs/1306.4447; Karan Ganju, Qi Wang, Wei Yang, Carl A. Gunter, and Nikita Borisov, "Property Inference Attacks on Fully Connected Neural Networks Using Permutation Invariant Representations," *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security,* CCS '18: 2018 ACM SIGSAC Conference on Computer and Communications Security, (Toronto Canada: ACM, 2018), 619-33, https://doi.org/10.1145/3243734.3243834; Yuantian Miao, Chao Chen, Lei Pan, Qing-Long Han, Jun Zhang, and Yang Xiang, "Machine Learning Based Cyber Attacks Targeting on Controlled Information: A Survey," *ArXiv:2102.07969 [Cs],* February 16, 2021, http://arxiv.org/abs/2102.07969.

[149] Matt Fredrikson, Somesh Jha, and Thomas Ristenpart, "Model Inversion Attacks That Exploit Confidence Information and Basic Countermeasures," *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security,* CCS'15: The 22nd ACM Conference on Computer and Communications Security, (Denver Colorado: ACM, 2015), 1322-33, https://doi.org/10.1145/2810103.2813677.

[150] Florian Tramèr, Fan Zhang, Ari Juels, Michael K. Reiter, and Thomas Ristenpart, "Stealing Machine Learning Models via Prediction APIs," 25th *USENIX Security Symposium (USENIX Security 16)* (Austin, TX: USENIX Association, 2016), 601-18, https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/tramer; Robert Nikolai Reith, Thomas Schneider, and Oleksandr Tkachenko, "Efficiently Stealing Your Machine Learning Models," *Proceedings of the 18th ACM Workshop on Privacy in the Electronic Society - WPES'19,* the 18th ACM Workshop, (London, United Kingdom: ACM Press, 2019), 198-210, https://doi.org/10.1145/3338498.3358646.

[151] Tramèr, et. al., "Stealing Machine Learning Models via Prediction APIs."

[152] Tegjyot Singh Sethi and Mehmed Kantardzic, "Data Driven Exploratory Attacks on Black Box Classifiers in Adversarial Domains," *Neurocomputing* 289 (May 10, 2018): 129-43, https://doi.org/10.1016/j.neucom.2018.02.007.

[153] Michał Choraś, Marek Pawlicki, and Rafał Kozik, "The Feasibility of Deep Learning Use for Adversarial Model Extraction in the Cybersecurity Domain," *Intelligent Data Engineering and Automated Learning – IDEAL 2019,* ed. Hujun Yin et al., Lecture Notes in Computer Science 11872, (Springer Cham International Publishing, 2019), 353-360, https://doi.org/10.1007/978-3-030-33617-2_36.

[154] Tramèr, et. al., "Stealing Machine Learning Models via Prediction APIs."

[155] Pak and Hwang, "Performance Improvement by Attribute Selection in the Network Intrusion Detection System."

[156] Ben Buchanan et al., "Automating Cyber Attacks," Center for Security and Emerging Technology, November 2020, https://doi.org/10.51593/2020CA002.

[157] Julien Salinas, "GPT-3 Open-Source Alternatives: GPT-Neo and GPT-J," NLP Cloud, July 14, 2021, https://nlpcloud.io/gpt-3-open-source-alternatives-gpt-j-gpt-neo.html.

[158] Adam Weidemann, "New Campaign Targeting Security Researchers," *Google Threat Analysis Group* (blog), January 25, 2021, https://blog.google/threat-analysis-group/new-campaign-targeting-security-researchers/.

[159] Ri Hyok Chol (리혁철), Kim Chol (김철), "A Method for Constructing Korean Spontaneous Spoken Language Corpus Based on Imitating *tho*(Particle)" (토의 모방에 기초한 조선어자연발화 음성언어코퍼스의 구축방법) Kim Il Sung University Gazette, 66, no. 3 (2020); Kim Tong Su (김동수), "A Method of Making of Tag Table with Appearance Frequency of Word in Corpus and Example Base Search Using It" ("코퍼스에서 단어의 출현빈도에 의한 꼬리표작성과 그것을 리용한 실례기지탐색의 한가지 방법"), *Kim Il Sung University Gazette,* 66, no. 1 (2020); Kim Chong Il (김청일), Ri Hyon Sun (리현순), "Study of the Recurrent Neural Network Language Model Using Future Context Information" ("미래문맥정보를 리용한 재귀신경망언어모형구축에 대한 연구"), Kim Il Sung University Gazette, 66, no. 3 (2020).

[160] Ben Buchanan, John Bansemer, Dakota Cary, Jack Lucas, and Micah Musser, "Automating Cyber Attacks," Center for Security and Emerging Technology, November 2020, https://doi.org/10.51593/2020CA002.

[161] RAND Interview #6, computer science and cybersecurity specialist, May 2021.

[162] Edward Geist and Andrew J. Lohn, *How Might Artificial Intelligence Affect the Risk of Nuclear War?* (Santa Monica, CA: RAND Corporation, 2018, PE-296-RC), 2, https://www.rand.org/content/dam/rand/pubs/perspectives/PE200/PE296/RAND_PE296.pdf.

[163] Michael Dahm, "Chinese Debates on the Military Utility of Artificial Intelligence," *War on the Rocks*, June 5, 2020, https://warontherocks.com/2020/06/chinese-debates-on-the-military-utility-of-artificial-intelligence/; John Costello and Joe McReynolds, *China's Strategic Support Force: A Force for a New Era* (Washington, DC: National Defense University Institute for National Strategic Studies, 2018), https://ndupress.ndu.edu/Portals/68/Documents/stratperspective/china/china-perspectives_13.pdf; Elsa B. Kania and John Costello. "Seizing the Commanding Heights: The PLA Strategic Support Force in Chinese Military Power," Journal of Strategic Studies 44, no. 2 (February 23, 2021): 218-64, https://doi.org/10.1080/01402390.2020.1747444.

[164] RAND Interview #1, North Korea expert, May 2021.

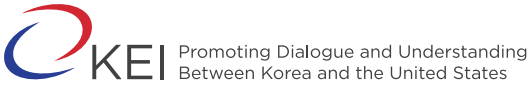
[165] RAND Interview #2, cybersecurity expert, May 2021.

[166] RAND Interview #5, cybersecurity expert, April 2021; Mun Dong Hui, "North Korea's Kumsong 121 Recently Employed Social Media to Launch a Cyber Attack," *Daily NK*, September 13, 2021, https://www.dailynk.com/english/north-korea-kumsong-121-recently-employed-social-media-launch-cyber-attack/#:~:text=North%20Korea's%20Kumsong%20121%20recently%20employed%20social%20media%20to%20launch%20a%20cyber%20attack,-The%20hacking%20group&text=The%20North%20Korean%20hacker%20group,cyber%20attack%20using%20social%20media-.&text=The%20attackers%20essentially%20grafted%20social,attacks%20aimed%20at%20particular%20individuals.

[167] RAND Interview #3, North Korea expert, April 2021.

[168] RAND Interview #5, cybersecurity expert, April 2021.

Contract Editor: Gimga Group | **Design:** Gimga Group



1800 K St. NW, Suite 300 | Washington, DC 20006
T 202.464.1982 | F 202.464.1987 | www.keia.org