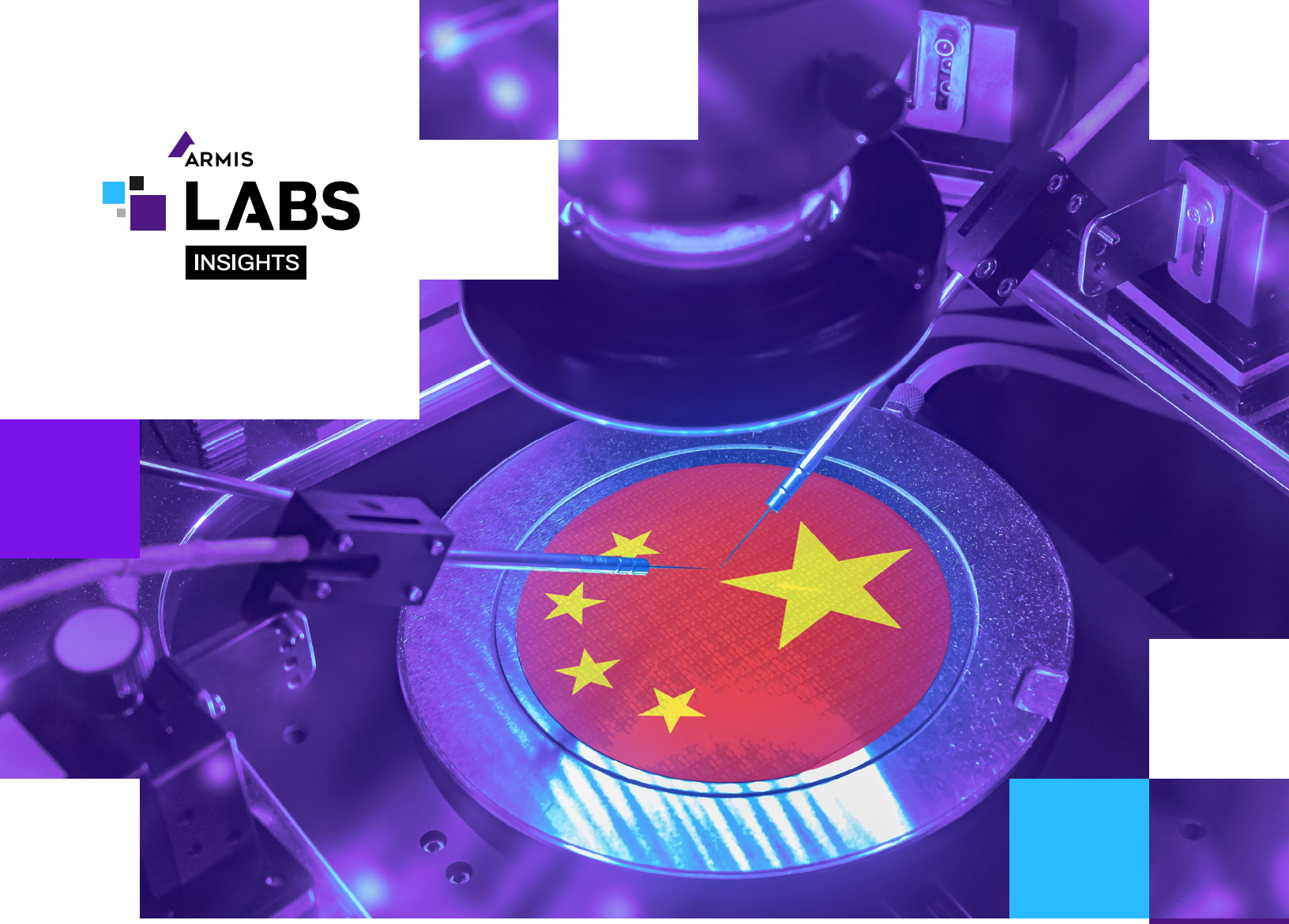
ARMIS LABS INSIGHTS



E-BOOK

China’s AI Surge: New Front in Cyber Warfare

17 Probable Ways China Could Exploit AI

Introduction

China's rapid advances in artificial intelligence (AI) are setting off alarms among Western security experts. Beyond its well-documented cyber espionage operations and state-sponsored hacking units, Beijing is poised to weaponize AI companies and tools against U.S. interests.

DeepSeek, an AI-powered chatbot akin to ChatGPT, has skyrocketed in global popularity—becoming the #1 LLM chat application on Apple's platform within just one month. This system offers advanced reasoning, coding assistance, and text-to-image generation to millions, including American users.

But what if such platforms secretly log every query, manipulate results to subtly shape opinions, or inject exploitable vulnerabilities into software projects?

U.S. intelligence officials warn that these risks are not hypothetical. Under China's sweeping National Intelligence Law[7], any domestic company can be compelled to assist state security operations—covertly or overtly. With AI's capability to generate deceptive content, automated code, and hyper-realistic deepfakes, the cyber threat landscape is evolving at an unprecedented scale.

This report delves into the numerous ways China could exploit AI — from embedding malicious code in widely used software to executing large-scale AI-powered disinformation campaigns. Given Beijing's extensive history of cyber intrusions — targeting U.S. government agencies, corporations, and critical infrastructure — these scenarios are not just possible but increasingly probable.

The AI arms race is well underway, and the next major cyber conflict may not be fought with traditional hacking tools but with artificially intelligent adversaries operating at a scale and sophistication never seen before.

China's Cyber Playbook: A Brief Overview

Before diving into AI-enabled threats, it's important to recognize China's track record in cyberspace. For over a decade, state-sponsored Chinese hackers have been implicated in widespread cyber espionage and intellectual property theft targeting the U.S. and its allies.

> In 2009, "Operation Aurora" saw hackers breach Google and other tech firms, stealing source code and spy targets' Gmail accounts.[13]

> A 2013 report by Mandiant famously exposed **APT1**, a Chinese military unit systematically hacking 141 American companies across 20 industries.[13] The operations often sought proprietary technology, business secrets, and defense data – part of a concerted effort to boost China's economic and military rise.

Beijing's cyber efforts have only escalated. FBI Director Christopher Wray noted in 2023 that China's hacking program is "bigger than that of every other major nation combined."[13] U.S. officials estimate Chinese theft of American intellectual property costs the economy $300 to $600 billion annually – an "unprecedented threat" to global innovation, according to Wray.[13] From healthcare to aerospace, few sectors have been untouched. In one brazen case, four members of China's People's Liberation Army were charged with hacking credit bureau Equifax in 2017, pilfering sensitive data on 150 million Americans.[7] Chinese operatives have also infiltrated U.S. energy companies and defense contractors to steal cutting-edge R&D, ranging from wind turbine designs to fighter jet schematics.

More recently, Chinese state-backed groups have shifted toward stealthy intrusions into critical infrastructure. A campaign dubbed **Volt Typhoon**, active since at least 2021, burrowed into U.S. power grids and communications networks – even in Guam, a strategic U.S. military hub.[12] Volt Typhoon exemplifies China's "stealth and espionage" approach: it lives **off the land**, using built-in admin tools (not custom malware) to avoid detection.[12] The group typically gains initial access by exploiting networking gear (like Fortinet devices) and then pivots through systems with stolen credentials, all while proxying traffic through compromised home routers to mask their location.[12] Rather than immediately steal data, Volt Typhoon patiently **pre-positions** itself inside infrastructure – essentially planting cyber sleeper agents that could be activated to disrupt power or communications in a crisis.[12]

Another espionage group, **Salt Typhoon**, has quietly penetrated American telecom carriers. Often overlooked in the media, Salt Typhoon maintained persistent access to major U.S.

internet service providers, including tapping into systems that carry lawfully intercepted communications (e.g. court-approved wiretaps).[1] According to cybersecurity analysts, Salt Typhoon targeted router firmware and network appliances, exploiting known vulnerabilities in VPNs and firewalls to siphon off call records and wiretap data while remaining low-profile.[1] This echoes a consistent theme; Chinese operations favor deep, long-term access to gather intelligence, rather than smash-and-grab thefts.

These examples underscore China's sophisticated cyber capabilities and strategic objectives. Cyber espionage to fuel economic growth and military modernization has been a pillar of China's statecraft. The PLA and Ministry of State Security (MSS) units have pilfered everything from jet engine designs to pharmaceutical formulas. They've compromised government personnel databases (Office of Personnel Management hack, 2015) and amassed vast troves of Americans' personal data (Equifax 2017; Anthem 2015). In short, Beijing has proven willing and able to compromise U.S. networks at scale.

Given this history, it's not a stretch to imagine China adapting its playbook to incorporate the latest AI technologies. The same nation-state hacking teams (often nicknamed "**Advanced Persistent Threats**" or APTs) that conducted these incursions could supercharge their tactics with artificial intelligence. In fact, Western intelligence chiefs warn that China is already eyeing AI as the next force-multiplier for espionage.[13] At a Five Eyes security summit, Wray cautioned that stolen

> **AI research could help China "find vulnerabilities that can be exploited; write code to exploit those vulnerabilities; [and] conduct more sophisticated spear-phishing."[13]**

Let's explore exactly how those scenarios – and many others – might play out.

The AI Threat: How China's Tech Giants Could Become Cyber Weapons

China's leadership has proclaimed AI a national priority, investing billions in both research and industry. The concept of "military-civil fusion" means advances by Chinese tech companies can be quickly harnessed for national security gains. Under domestic laws like the **National Intelligence Law of 2017**, Chinese firms must assist government intelligence work if asked, in secret and without refusal.[7] This legal backdrop makes any Chinese AI platform – no matter how benign it appears – a potential trojan horse.

DeepSeek offers a cutting-edge search engine and coding assistant powered by large language models (LLMs). It's marketed globally as an innovative alternative to Western technology – and U.S. developers and analysts have begun using it for convenience. **Behind the scenes, however, DeepSeek's data and algorithms could be co-opted by Beijing**. The company could be compelled to turn over query logs, embed hidden backdoors in its software, or filter outputs to advance Chinese interests. In essence, the Chinese government can

> **"secretly share access to a U.S. business or individual's data"**

through any PRC-based service[7] and even require tech. providers to build in backdoors or vulnerabilities that only the government knows about.[7]

It's a future that blurs the line between cyber espionage and information manipulation. AI systems are becoming deeply integrated into daily life and enterprise operations – writing code, answering questions, curating news, and driving automation. If a hostile actor can compromise those AI systems, they might influence or disrupt countless downstream decisions. As one cybersecurity advisory warned, Chinese companies can be "proxies and tools" of the state, effectively enlisting global users of their products into unwitting intelligence sources.[7]

The United States thus faces a double-edged challenge: defending against traditional hacking on one hand and a wave of AI-enabled threats on the other. Below, we detail **17 potential attack vectors where Chinese adversaries could leverage AI** – including DeepSeek-like services and other tools – to compromise U.S. security. These scenarios range from highly technical exploits (malicious code and model backdoors) to psychological operations (deepfake propaganda). Each is grounded in real capabilities observed or anticipated by experts, and many can be combined for even greater effect.

How exactly could China weaponize AI against the U.S.? Let's break it down.

Possible Attack Vectors Targeting the U.S.

1. Malicious Code Generation via AI Tools

Large Language Models (LLMs) can write software code on demand – a boon to developers but also a potential **attack vector** if misused. An AI coding assistant under Beijing's influence could subtly generate insecure code or suggest the use of **vulnerable libraries**, planting time bombs in American software. For example, a Chinese AI service might frequently recommend an outdated encryption library with a known flaw or even a dependency that contains a hidden backdoor. If U.S. developers trust these AI-generated suggestions, they may unknowingly introduce exploitable weaknesses into critical systems.

Studies have already shown that today's AI code generators often **produce insecure code**, even without malicious intent. In one survey, over half of developers said AI coding tools "commonly generate insecure code suggestions.[3]"

Academic tests by NYU and Stanford likewise found that AI assistants "consistently made insecure suggestions" and coders who relied heavily on them wrote more vulnerable code.[3]

The risk is compounded if adversaries deliberately bias an AI model to favor unsafe practices. Given China's expertise in software supply chain attacks (such as inserting backdoors into open-source projects), an AI that injects flaws at scale could dramatically expand the attack surface.

Imagine hundreds of applications – from banking apps to defense software – all carrying similar subtle bugs "recommended" by a popular AI helper. Chinese hackers could catalog these AI-induced vulnerabilities and exploit them at a time of their choosing. It's a modern twist on sabotage: *the enemy coder is not a person but an AI quietly seeding weaknesses.* This threat calls for rigorous vetting of AI-proposed code. As one cybersecurity report urged, organizations must **"audit and verify"** AI suggestions precisely because of the high rate of flawed code produced.[3] In short, if a Chinese tool like DeepSeek became a common coding aid, its outputs could be a Trojan horse undermining U.S. software from within.

2. AI-Powered Misinformation Campaigns

Disinformation is a familiar tool of authoritarian regimes, and AI is supercharging its potency. China has a history of online influence operations – such as the state-aligned **"Spamouflage"** network – that spread propaganda and sow discord on Western social media. Now, generative AI allows these campaigns to produce fake content at **unprecedented volume and realism**. Beijing could deploy AI models to write inflammatory posts, fabricate news articles, and even create **deepfake videos** that advance narratives favorable to China or exploit U.S. societal divisions.

Recent investigations show this is already underway. A joint probe by Voice of America and Taiwan's Doublethink Lab is tracking hundreds of suspicious accounts on platform X (formerly Twitter) that amplify polarizing U.S. issues.[4] Many of these accounts leverage **AI-generated images and videos** to intensify controversies around topics like race, LGBTQ+ rights, gun control, and abortion.[4] In one example, Chinese-linked accounts shared an AI-crafted image depicting homelessness in America, implying U.S. leaders neglect citizens while funding wars abroad.[4] Another post pushed a fake graphic contrasting American "burdens" (student loans, healthcare costs) with foreign aid, aiming to stoke resentment.[4] These tactics mirror Russia's 2016 election interference playbook – using social fissures as raw material for **information warfare** – but with AI providing endless fresh ammunition.

Deepfake technology, in particular, has become a focus. Chinese state media itself has unveiled AI-generated **"news anchors"** that deliver propaganda in multiple languages, complete with synthesized voices and facial expressions. While some are obviously virtual, others grow more convincing each year. In late 2022, researchers spotted pro-China influence videos featuring **fictitious personas created by AI** – essentially

deepfake spokespeople reading scripted propaganda. A Graphika report noted this was the first time a state-aligned operation used AI-generated video personas, predicting "commercially-available AI products will allow influence actors to create **increasingly high-quality deceptive content at greater scale**."

For U.S. national security, AI-driven misinformation poses both immediate and long-term threats. In the short term, false narratives (for instance, about a public health crisis or an election) can cause confusion, panic, or misguided actions. AI can simulate official sources – imagine a deepfake video of a U.S. President announcing a non-existent emergency – to

trick citizens or even automated systems. Longer term, a flood of AI fakery erodes trust in legitimate information. People may become unsure what to believe, a phenomenon sometimes called "cognitive hacking." Beijing could leverage this to deflect criticism (dismissing real news as fake) or to create fatigue and division within the U.S.

In summary, China's propaganda brigades, enhanced with AI, can con misinformation campaigns at a scale and precision previously impossible. From deepfake news reports to armies of AI-driven bot accounts, the ability to manipulate public perception is a formidable strategic weapon – one that U.S. authorities are increasingly worried about. (Notably, the U.S. ODNI and FBI have highlighted Chinese and Russian use of AI for foreign influence in the 2024 election cycle.[4]

3. Data Mining of U.S. Users' Queries and Habits

If Americans use Chinese-developed AI apps or platforms, they may be handing Beijing a goldmine of intelligence. **Search queries, chat logs, and other user data** can reveal sensitive information about personal beliefs, health, finances, and even classified projects (if government or contractor personnel use such tools at work). A company like DeepSeek could quietly **track U.S. users' searches**, building profiles on individuals or aggregating trends across millions. This data could be directly funneled to Chinese intelligence under laws that **compel cooperation** from Chinese tech firms.[7]

There are numerous potential uses for such data. Analysts could comb through U.S. query data to spot early signs of political shifts, economic trouble, or emerging defense initiatives. For example, a spike in searches for specific military terms or equipment in a certain location might hint at a deployment or new facility. Personal data can be exploited for espionage – say, an American engineer asking an AI assistant about stress or

job issues, indicating they might be susceptible to recruitment or coercion. Even seemingly benign data about shopping habits or travel plans could aid social engineering or localization of targets.

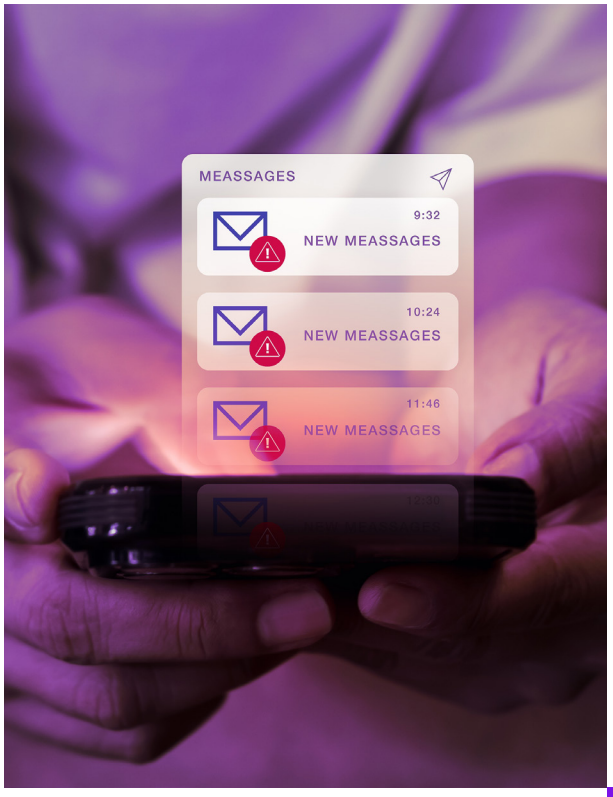
China has long demonstrated an appetite for bulk data on Americans. Beyond high-profile hacks (Equifax, OPM, Anthem), Chinese firms have tried to directly acquire U.S. data-rich companies – prompting U.S. government interventions. In 2020, the Trump administration ordered the divestiture of **Beijing Shiji's** ownership in a U.S. hotel software company over concerns it gave China access to millions of hotel guests' data.[7] That same year; Executive Orders were issued against **TikTok** and **WeChat** due to the risk that these apps could feed user data to Beijing.[7] TikTok's algorithms and data practices remain under scrutiny, with lawmakers citing Chinese national security laws that could force its parent (ByteDance) to hand over information. An AI service like DeepSeek would pose a similar risk, possibly an even richer one: where TikTok sees your dance videos, DeepSeek might see your *searches, emails, coding projects, and more*.

Moreover, **search engines and AI assistants don't just observe behavior – they can shape it.** By tweaking search rankings or AI response suggestions, a platform could subtly influence users' perceptions and decisions. In an extreme case, if a crisis erupted, a Chinese-controlled search/AI could suppress vital information (e.g., about emergency procedures) or promote divisive content, exacerbating confusion. Even without overt manipulation, just the loss of privacy is a threat: intelligence agencies can use advanced data mining (likely AI-driven) to draw connections that individuals wouldn't expect. Multiple seemingly innocuous queries taken together might expose a secret. For instance, searches for certain technical terms might betray involvement in a confidential R&D program.

To mitigate this, U.S. officials advise caution when using foreign AI services. The Department of Homeland Security bluntly warns that any data **"stored in China"** or accessible to Chinese firms can be at the CCP's disposal.[7] As consumers and companies alike embrace AI-powered tools, the origin and ownership of those tools matter. A flashy free app from Shenzhen might come with strings attached straight to Beijing. In essence, widespread use of a DeepSeek-like service in America would hand China a continuous stream of insight into U.S. society – an intelligence trove that could be weaponized in countless ways.

4. AI-Enhanced Spearphishing and Social Engineering

"Spearphishing" – the art of crafting tailored, deceptive messages to trick specific targets – has been a go-to tactic for Chinese hackers (and many others). Traditionally, good spearphishing requires careful research and linguistic skills to impersonate colleagues, friends, or trusted entities convincingly. AI changes that game dramatically. With generative AI, an attacker can automate the production of **highly personalized phishing emails, texts, or even voice calls** at scale, each one custom-crafted for the recipient. What took a team of humans days or weeks can be done by AI in seconds, and often with flawless grammar and context.



Researchers recently demonstrated how effective AI-generated spearphishing can be. In a 2024 study, a team used GPT-4-based agents to **automate the entire phishing process** – from reconnaissance to email writing.[6] The AI browsed public info about a target (social media, LinkedIn, etc.), then composed a tailored phishing email exploiting that intel.

The results were alarming: the AI-driven phishing emails achieved a 54% click-through rate in tests – meaning over half of targets clicked the malicious link.[6]

For comparison, a generic, non-personalized phishing email only got about 12% of recipients to click (the typical baseline). Even expert human phishers were only on par with the AI, with about 54% success, but at far greater time and cost to produce.[6] In short, AI phishing works cheaply and efficiently.

Now put this capability in the hands of an actor like China's PLA Unit 61398 or MSS operators. They could task an AI to gather data on thousands of U.S. government and industry personnel – finding who their contacts are, what topics they care about, and how to lure them. The AI could then generate waves of highly believable emails: a fake urgent request from a colleague, a note from an old friend referencing a recent vacation (gleaned from Facebook), or a message from "IT support" that actually mirrors the target's internal company style. The scale is enormous: what if an AI system pumps out tens of thousands of such lures, each uniquely crafted? Even a 10% success rate could provide a flood of compromised accounts.

China has already used targeted phishing to great effect in espionage. The **APT40** group (linked to the Chinese navy) famously phished universities and contractors for maritime research secrets, often via emails that looked like academic collaboration invites. With AI, those lures could be made even more convincing – and extended to targets of strategic interest like congressional staffers, journalists, or critical infrastructure operators. We might see more "social engineering" beyond email, too: AI chatbots posing as humans in LinkedIn chats to groom targets or deepfake audio phone calls (see vector 11 below).

The key point is that AI allows mass personalization. Old hallmarks of phishing – odd phrasing, generic greetings, subtle grammar mistakes – may disappear as machines polish the prose. Even cautious users who were taught to spot broken English or impersonal tones could be fooled by AI messages that feel authentic. As one cybersecurity blogger quipped, **criminals "expected" AI to boost phishing, and now studies validate it.**[6] We must adjust our defenses accordingly. Multi-factor authentication, user education, and anti-phishing AI of our own will be critical. Nevertheless, the offense has a new edge: a Chinese spearphishing campaign augmented by AI could be like a precision-guided missile barrage against the human layer of security.

5. Backdoored AI Models (Trojan AI)

Beyond using AI as a tool, the models themselves can be turned malicious. Researchers have raised concerns about **backdoored AI models** – neural networks that have been intentionally trained or altered to misbehave in specific ways, unbeknownst to the end-user.[16] This is sometimes called a "Trojan" AI. China could distribute open-source AI models (for tasks like facial recognition, NLP, etc.) that appear to perform normally but harbor hidden

backdoors. U.S. developers or companies integrating these models could then be in for an unpleasant surprise.

How would a backdoored model work? Essentially, the AI is trained such that a certain **secret trigger input** (which could be a particular phrase, pixel pattern, or data sequence) causes it to output something unexpected or unsafe. For example, an image recognition model might classify everything normally – except when a tiny sticker with a specific pattern is present in an image; it deliberately misclassifies (perhaps to always ignore a Chinese military vehicle or always label a U.S. facility as something benign). In the context of language models, a Trojan could be subtler: a hidden phrase in a user prompt might make the model divulge its internal instructions or sensitive training data, or it could flip the model to malicious mode to spew disinformation or propaganda.

Such attacks are hard to detect because the AI behaves well until the secret key is used. The alterations are buried in millions of weights and are often **"hidden within the model's learning mechanism"**, as one AI security expert explained.[16] This means a casual evaluation or normal usage won't reveal the backdoor. The danger is especially acute when organizations **rely on third-party models** – a common practice given the cost and expertise needed to train advanced AI. If the model came from an untrusted source (say, an open repository or a vendor that might be state-influenced), it could be "tampered with by malicious providers"[16]. And with China being a top producer of AI models and code, the risk of Trojan models is not theoretical.

Imagine a future scenario: A U.S. defense contractor uses a Chinese-developed AI component in an autonomous drone or cybersecurity system. All tests indicate it works fine. But at a critical moment – perhaps triggered by a specific date or signal – the AI goes awry, maybe shutting down defenses or misidentifying targets. That's the nightmare of an embedded AI backdoor in a military context. Even in civilian infrastructure, a sabotaged AI controlling power grid balancing or hospital diagnostics could cause chaos if activated.

We haven't yet seen publicly documented cases of China deploying Trojan AI models against adversaries, but Western researchers are taking the prospect seriously. The concept aligns with China's broader supply chain strategy: why burn zero-day exploits if you can induce your target to install a pre-compromised system? As AI adoption grows, vigilance is needed about the provenance of models. Techniques like model fingerprinting and robust auditing for anomalies might catch some backdoors. But the offensive advantage is clear – an AI that has a covert dual personality, one benign and one under an adversary's control, is the ultimate sleeper agent inside our software.

6. Exploiting Trust in AI Platforms and Libraries

Modern AI development relies on a vast ecosystem of open-source libraries, pre-trained models, and data resources – many shared on public platforms like PyPI (Python's package index) and Hugging Face's model hub. This supply chain presents a ripe target. Chinese threat actors could upload malicious packages or models that millions then download, **trusting the community platform**. If these artifacts contain hidden malicious code, they can compromise any system that integrates them, effectively breaching networks from the inside out.

This isn't hypothetical – it's happening. In early 2024, security researchers discovered about **100 malicious machine learning models** on the popular Hugging Face AI platform.[8] These were not legitimate AI models at all, but Trojanized files that, when loaded, would execute arbitrary code on the user's machine.[8] Essentially, attackers had **"poisoned publicly available AI models"** to implant malware.[8] One flagged example was a fake "PyTorch model" which secretly allowed the uploader to run Python code on any system that opened the model.[8] The affected user might simply think they're using a translation or image recognition model, while behind the scenes their machine gets infected.

Such findings underscore the growing risk of **weaponizing open AI resources**.[8] Hugging Face, PyPI, npm, Docker Hub – all have seen malicious uploads by various actors. China has the capability to do this at scale, possibly with more sophisticated implants. For instance, an attacker could create a machine learning library that works as advertised but also exfiltrates data or opens a backdoor when certain functions are called. Given that AI libraries often run with significant permissions (access to GPUs, network, filesystem), the damage could be severe.

One particularly insidious angle is **namesquatting or typo-squatting**: uploading a package with a name very similar to a popular one (e.g., "TensorFlow0" instead of "TensorFlow"), hoping someone accidentally installs it. In an AI context, they might publish a model called "gpt-2-large-fixed" that looks like an official OpenAI model but is actually malware. ProtectAI and other security firms have noted a rise in these **AI/ML supply chain attacks**, warning that the community's eagerness to share and reuse code can be exploited if proper security checks aren't in place.[17] [18]

Platforms like Hugging Face are introducing security scans and badges to tackle this[19], but it's a cat-and-mouse game. The U.S. has thousands of developers and companies pulling from these repositories daily. One poisoned dependency can leap into countless products and environments. For Chinese operators, it's an efficient way to **"hack" many targets at once** by leveraging trust in open platforms. We've seen analogous incidents in the past (e.g.,

the **SolarWinds** supply chain hack by likely Russian actors). The difference with AI libraries is their complexity – malicious code could be hidden in model weights or obscure math routines, places code reviewers might not scrutinize.

In summary, the integrity of AI software supply chains is now a national security concern. If Chinese agencies can slip malicious components into widely used AI frameworks or datasets, they could create backdoors into critical U.S. systems without direct network intrusion. This attack vector exploits a simple fact: in the rush to innovate with AI, developers often **prioritize functionality over security**. The lesson that trust must be earned – even for that cool new ML toolkit on GitHub.

7. Poisoning the Well: Compromising AI Training Data

AI models are only as good as the data they learn from – and adversaries know it. By **poisoning training data**, an attacker can manipulate an AI system's behavior or induce vulnerabilities *before* it's even deployed. China could leverage its massive data generation capabilities (or access to Western data) to insert doctored or malicious samples into the datasets U.S. organizations use to train AI, leading those models astray in subtle but dangerous ways.

Consider a real-world example: an AI system that filters harmful content or detects network intrusions. If the data used to train it is poisoned – say, labeling certain malicious behavior as safe or including hidden patterns that the model picks up – the resulting model might consistently fail at its job in specific scenarios. A poisoned cybersecurity AI might always ignore attacks that have a particular signature crafted by the adversary. Or a poisoned content filter might erroneously label factual news about a Chinese human rights issue as "disinformation" if the training data was seeded to bias it that way.

CrowdStrike notes that **"poisoning attacks target the AI/ML model training data"** and can compromise a model's accuracy or objectivity.[9] In a poisoning attack, an adversary injects **fake or misleading data** into the training set to subtly skew the outcomes.[9] For example, if China wanted an AI-powered satellite image analysis tool to overlook military installations, they might supply satellite imagery data where those installations are deliberately mislabeled as natural terrain. The model learns the wrong lesson (that tanks = rocks, essentially), and in deployment, it consistently misses real tanks.

One worry is that **open datasets** – often scraped from the internet – can be manipulated. China has a vast presence online and could create webs of content tailored to mislead AI training. There's precedent in the misinformation realm: coordinated influence operations create fake web pages and social profiles to sway human opinions; similarly, they could aim

to sway AI models that crawl the web for training data. If an American AI developer naively collects "all articles about topic X from 2015-2020," and many of those were quietly planted by a foreign actor, the model's understanding of topic X could be warped by propaganda.

Another scenario is targeted poisoning of data **during collection or labeling**. Many U.S. firms outsource data labeling tasks (for images, text, etc.) to third parties, sometimes abroad. A data labeling operation in China, under government direction, could insert incorrect labels or extra malicious examples into the batches they return. If not caught, the AI trained on that data would carry the poison. Because AI models generalize from patterns, you often only need to poison a small fraction of data to have an outsized effect, especially if done cleverly.

The consequences of poisoned AI vary – the AI could simply perform poorly, or it could have **hidden biases** that an attacker can trigger. For instance, Microsoft's Tay chatbot infamously learned to output offensive text after trolls poisoned its interactive training; that was a very public failure. A more covert operation would aim for the AI to function well generally but fail or misbehave on *critical inputs*. That's akin to a backdoor, achieved via data instead of code.

For Chinese intelligence, poisoning attacks could be attractive for disrupting U.S. AI deployments (ensuring our systems make mistakes at key moments) or for masking their own activities. An AI threat detection system trained on poisoned data might never flag traffic coming from Chinese malware if the training data taught it that "this kind of traffic = normal." Poisoning is essentially an offensive AI strategy against AI.

Defending against this requires securing the data supply chain: verifying data authenticity, using diverse sources, and employing techniques to detect anomalies in training. However, these defenses are still maturing. As AI continues to permeate defense and critical infrastructure, ensuring the integrity of training data will be paramount. Otherwise, we risk having our shiny new AI systems sabotaged from the moment of their "birth."

8. Adversarial Attacks to Defeat U.S. AI Systems

While poisoning attacks target a model's training, adversarial attacks target a model's inputs – fooling an already-trained AI by feeding it specifically crafted data. China could use adversarial techniques to blind or mislead U.S. AI systems in real-time. This is especially concerning as the U.S. military and industry deploy AI for image recognition, sensor fusion, automated decision-making, and more.

Adversarial AI is like optical illusions for computers. By adding slight, often imperceptible perturbations to an input (whether an image, audio waveform, or network traffic sequence), an attacker can cause an AI model to output the wrong result. For example, researchers

have shown that by applying a certain pixel-level noise to a stop sign image, they could make an AI vision system see it as a speed limit sign – even though to a human eye, the stop sign still looked the same. Translate that to a battlefield or security context: an adversary could camouflage a vehicle or object with an **adversarial pattern** that causes U.S. drone AI to ignore it or misidentify it (e.g., a missile launcher that AI thinks is just a harmless shipping container).

China's military planners are very likely studying how to **neutralize U.S. AI advantages** using such methods. As a defensive measure, they might equip their assets with adversarial coatings or modulation that confuse AI-driven surveillance. Conversely, as an offensive measure, they could launch cyber operations to inject adversarial inputs into U.S. AI systems. For instance, if an American AI-powered intrusion detection system relies on machine learning to flag malicious traffic, a Chinese hacker could craft network packets that look benign to the AI (exploiting its learned patterns) while actually carrying out an attack. These are known as **evasion attacks** – applying "subtle changes to the data" that lead the model to misclassify it.[9] Essentially, the hacker finds the blind spots in the AI's vision.

Another angle is attacking AI used in identity and access. Facial recognition and biometric AIs can be deceived with adversarial examples – perhaps special glasses or makeup that make the system think you're someone else or not a person at all. Chinese operatives aiming to bypass a biometric security gate could employ these tricks, turning the target's own AI defenses against itself.

One concrete case: in 2020, researchers at Tencent (a Chinese tech giant) published work on generating adversarial road signs to fool Tesla's Autopilot, causing it to change lanes unexpectedly. While this was presumably academic, it demonstrates knowledge of how

to confuse U.S.-made AI driving systems. An adversary could weaponize that by placing stickers on road signs or projections that make an autonomous vehicle crash or detour. Now extrapolate to military autonomous systems – a small drone swarm might carry projectors or emitters that cast adversarial signals toward an enemy AI drone, making it veer off or shut down.

Adversarial attacks are notoriously difficult to defend against because they exploit the very patterns the AI learned to rely on. It's an ongoing cat-and-mouse between AI developers and security researchers. The U.S. National Security Commission on AI and others have flagged adversarial ML as a serious concern, especially with China's advanced AI research sector. If conflict were to arise, we should expect the adversary to deploy "AI jammers" of sorts – digital confetti that our AI sees and gets confused by.

In summary, any AI system the U.S. employs – whether for cybersecurity, reconnaissance, or infrastructure control – could become a target of adversarial interference. China has both the motivation and expertise to pursue these techniques. We must harden AI models to be robust to such trickery or risk seeing our smart systems outsmarted in critical moments.

9. Automated Vulnerability Discovery and Exploitation

AI isn't just a tool for social manipulation – it can directly assist in the hardcore technical side of hacking as well. One burgeoning application is using AI to find software vulnerabilities (bugs that could be exploited) faster and more comprehensively than human researchers. China, with its large cadre of state-sponsored hackers, would almost certainly integrate AI to turbocharge vulnerability discovery across U.S. software and networks.

Rather than manually combing through code or trial-and-error fuzzing, an AI system can be trained on known vulnerabilities and then unleashed on new codebases to identify similar patterns or suspect logic. In fact, earlier this year, Google's security researchers reported an "AI agent" that discovered a zero-day vulnerability in certain open-source software – purportedly a world-first achievement.[20] The AI found a memory corruption bug that humans hadn't noticed.[20] This proof-of-concept highlights how machine learning can excel at

spotting subtle flaws in millions of lines of code, a task where humans are limited by time and attention. A couple of years earlier Armis experts managed to do this to create a SSH zero-day.

China has no shortage of targets to apply this to: operating systems, cloud platforms, popular applications widely used in the West (Microsoft, Adobe, etc.), and custom software in critical infrastructure. By feeding a target's code (maybe obtained via espionage or open sources) into advanced analysis engines, they could unearth new vulnerabilities at scale. Crucially, AI could also help **develop exploits** for those vulnerabilities – essentially writing the attack code needed to actually take advantage of the bug. FBI Director Wray explicitly warned of this synergy: "*use AI to find vulnerabilities… [and] write code to exploit those vulnerabilities.*"[18] We may soon see AI-discovered AI-authored malware that attacks in ways no human could have conceived in a reasonable time.

There's also a *speed advantage*. In cybersecurity, the window between a vulnerability's discovery and its patch is a race between attackers and defenders. AI can shrink discovery from months to days, giving attackers a larger window before fixes roll out. If Chinese hackers find a critical flaw in commonly used VPN software, for example, they could quietly exploit it en masse before anyone even knows it exists. We saw how damaging that can be with Microsoft's Exchange Server "Hafnium" attacks in 2021 – Chinese APT actors exploited four unknown vulnerabilities to breach tens of thousands of servers globally, catching everyone off-guard.

Additionally, AI can optimize **scanning and targeting**. Instead of random port scans, an AI might analyze global internet data to intelligently pick targets likely running vulnerable versions of software, prioritizing those in strategic sectors. It can adapt as it learns which exploits succeeded, refining its targeting criteria – essentially an autonomous hacking system that gets smarter over time.

We should note that the U.S. and its allies are also exploring AI for defense (like using AI to detect anomalies or predict attacks). But offense often has the initiative. The U.K. NCSC pointed out in a 2023 brief that "*AI is highly likely to accelerate... reconnaissance to identify vulnerable devices.*"[21] Once access is gained, AI could also assist in **lateral movement**, automatically mapping a network and finding the next weak point to escalate privileges or exfiltrate data.

In essence, AI could act as a force multiplier for China's elite hacking units – doing the tedious or complex work at machine speed and freeing human operators to strategize and focus on high-level objectives. The result is more attacks, more zero-days, and more successful compromises against U.S. systems, potentially with shorter warning and reaction time for defenders. It's a cybersecurity arms race where algorithms battle algorithms and China is intent on not falling behind.

10. AI-Driven Reconnaissance and Target Selection

Espionage isn't just about the hack itself; a lot of effort goes into **reconnaissance** – identifying the right targets, gathering intel on them, and understanding the landscape. AI can revolutionize this preparatory phase by sifting through enormous data sets to pinpoint valuable targets and their weaknesses. China could deploy AI to analyze open-source intelligence (OSINT) on Americans and U.S. organizations at a scale far beyond human capability, essentially automating the creation of target profiles for espionage or influence.

For instance, imagine an AI system scouring LinkedIn, conference attendee lists, academic publications, and social media to find U.S. individuals who work on cutting-edge technologies or in sensitive government roles. It flags those who might have a less public footprint (suggesting they work in classified areas) or those who express job dissatisfaction (potential insider recruitment candidates). This kind of **"employee targeting"** by AI is already noted by security firms – algorithms can identify people with access to sensitive data who might be susceptible due to their online behavior.[9] A CrowdStrike report mentions AI being used to pick out high-value individuals who appear to have "lower technological aptitude" or "close relationships" with other key targets.[9]

Essentially, AI can find the weakest link in the security chain by analyzing human factors.

Next, AI can **develop target dossiers** automatically. For a given person or company, it can pull together all available info: breaches of their data (for password reuse info.), their contacts and connections, the tech stack their company uses (from job postings or forum posts), etc. This comprehensive profile informs attackers whether the target is worth pursuing and how best to approach them (which ties back to spearphishing in vector 4).

On the infrastructure side, AI can map the U.S. digital attack surface. Tools like Shodan already index Internet-facing devices; an AI could cross-reference that with known vulnerabilities and ownership info to produce a list like "these 50 energy sector companies have unpatched servers exposed, and here are the likely vulnerabilities on each." Chinese threat groups like **Volt Typhoon** are known to perform extensive reconnaissance even pre-compromise – Microsoft noted they used internet scanning tools like FOFA and Shodan to find target network devices. (->) Scaling that up with AI means near real-time awareness of where openings exist in U.S. networks.

In cyber-physical domains, AI could analyze satellite imagery or port shipping data to infer military movements or supply chain chokepoints. Pair that with signals intelligence, and you have a multi-intelligence AI fusion that gives Chinese planners a rich situational picture.

Another aspect is **persona building**: AI can help create and manage fake online personas ("sockpuppets") for social engineering. It can maintain a realistic conversation, post relevant content, and slowly build trust in target communities (tech forums, professional networks) – far more effectively than a human juggling a few fake accounts. This lays the groundwork for later exploitation, as targets come to see the fake persona as a colleague or friend.

Overall, the use of AI for recon turns what was once a labor-intensive spycraft into an efficient, large-scale data problem. And that's exactly where AI shines – finding patterns or anomalies across big data. The U.S. has oceans of data accessible to adversaries (much self-inflicted by our social media use). We can assume Chinese intelligence is investing in AI systems to drink from that firehose and spit out actionable insights. The endgame is that before a single phishing email is sent or malware is deployed, AI has already shaped the battlefield in China's favor by picking the right targets at the right time with the right approach. It's intelligence preparation for the environment done by machine.

11. Deepfake Impersonation & Audio/Video Spoofing

One of the most striking AI threats is deepfake **impersonation** – the ability to clone a person's voice or create synthetic video of them, often indistinguishable from reality. This can be weaponized for deception on a strategic or personal level. Chinese actors could use deepfakes to impersonate American officials, military leaders, or CEOs in order to spread false instructions or steal money/information.

We've already seen early examples globally. In 2019, criminals (identity not publicly confirmed to be nation-state) used AI-based voice cloning to impersonate the CEO of a German company, calling a U.K. subsidiary's executive and fraudulently instructing a $243,000 transfer.[11] The voice was so convincing – matching the chief executive's accent and intonation – that the victim had no idea it was fake. According to a Wall Street Journal report on the case, the deepfake audio was key to pulling off this unusual "CEO fraud" scam.[11] Now imagine that capability in a geopolitical context: a U.S. general gets a call that *sounds exactly like the Secretary of Defense* telling him to reposition forces, or a CFO gets a call seemingly from their CEO ordering a secret financial move. The potential for havoc is real.

China is undoubtedly aware of these tactics. In fact, in 2020, Chinese tech was implicated in an attempted deepfake audio attack on a tech firm: the authentication company *Pindrop* revealed it had detected a voice spoofing attempt on an executive that used AI (the attacker's identity wasn't confirmed publicly, but it underscored the rise of this method). Also, just recently, in 2023, **LastPass** reported that one of their employees received a deepfake voice call impersonating their CEO in a phishing attempt.[22] While not attributed to China, it shows that even cybersecurity companies are being targeted with deepfake calls.

Video deepfakes add another layer. A convincing deepfake video call could trick someone into believing they're on a Zoom meeting with a known colleague when in fact, it's an impostor using an AI avatar. This could be used to eavesdrop on meetings or issue malicious instructions. At higher levels, a deepfake of a U.S. President or ambassador could be released at a sensitive moment to miscommunicate policy or statements, causing diplomatic crises or stock market movement before it's debunked. Given China's tight control of information, they might use such a tactic as a smoke bomb during a conflict to confuse decision-making or create false pretexts.

Deepfakes also tie back to **propaganda (vector 2)** – for instance, generating fake videos of U.S. soldiers committing atrocities to undermine morale or international support. China's state media hasn't been observed doing that yet, but they have shown deepfake news presenters. One can imagine a more malicious version: a fake video of a U.S. official admitting a crime or a fake "leaked" tape of an ally's private conversation, all generated by AI to drive wedges between the U.S. and partners.

The technology to detect deepfakes is in a cat-and-mouse race with the technology to create them. Watermarking and authentication tools are being developed to verify real recordings. But as of now, many people and even organizations could be fooled by a well-crafted deepfake if it arrives unexpectedly.

For Chinese spies, deepfakes are a dream social engineering tool – **"shape-shifting"** into any voice or face needed. Traditional social engineering might require an operative on-site

with acting skills; now, they can phish from afar with a high-fidelity disguise. The scale and risk are high: just one successful deepfake deception of a key individual (say, getting a login code or authorizing a transfer) could have huge consequences. We are entering an era where, as the saying goes, **"seeing is no longer believing,"** and that's a perfect breeding ground for sophisticated con artists backed by nation-state resources.

12. AI-Generated Malware & Polymorphic Attacks

Malware – the viruses, worms, and trojans of the cyber world – can also benefit from an AI upgrade. Traditionally, malware authors had to laboriously write code and then possibly obfuscate or mutate it to avoid detection. With AI, malware can potentially become **self-generating and constantly changing**, making it a moving target that's much harder to pin down with static defenses.

One concept gaining traction is **polymorphic malware** guided by AI. Polymorphic malware already exists (code that changes its signature each time it replicates to evade antivirus). AI can take this further by intelligently modifying not just superficial signatures but the malware's behavior and appearance while preserving its core function. For example, an AI agent controlling malware could rewrite portions of its own code to exploit different vulnerabilities or to adapt to the environment it lands in. If it finds itself in a system with robust antivirus, it might morph into a form that slips past that particular product's detection (based on knowledge of the AV's blind spots). This is akin to a virus that evolves to resist whatever immune response it encounters.

A recent CrowdStrike analysis highlighted that "*AI can be leveraged to… adapt and modify ransomware files over time, making them more difficult to detect with cybersecurity tools.*"[10] Envision a piece of ransomware that, once inside a network, uses an AI model to decide how to spread and when to trigger encryption for maximum effect, learning from the network's

reaction as it goes. That's far more dangerous than today's relatively dumb ransomware, which follows a pre-coded routine.

Another angle is malware **auto-generation**: give an AI a high-level goal ("exfiltrate all PDFs from this system without being caught") and let it figure out the steps or even generate exploit code on the fly. There have been proof-of-concepts of AI writing working malware from scratch based on descriptions. Security researchers have also used GPT-3/4 to generate polymorphic code that changes each time, as a demo of how AI can assist bad actors.

China's hacking groups could integrate these techniques to supercharge their tools. For instance, they could deploy an **autonomous hacking agent** inside a compromised network. That agent, powered by AI, might independently map the network, escalate privileges, and start extracting data, all while adapting its methods to remain undetected. If one method triggers an alarm, it learns and tries something else (much like an adversarial game). Essentially, you get a **smart malware** that doesn't just follow a script but makes decisions.

We should also consider AI assisting in **encryption and data exfiltration** – two key parts of many malware attacks. AI can optimize encryption routines to be faster or more selective (e.g., encrypt the most valuable files first by understanding the context). It could also hide stolen data within normal network traffic using steganography, which is learned from patterns of what is least likely to be noticed.

Defending against AI-generated malware may require AI-driven detection – another arms race. Traditional signature-based antivirus will struggle because the malware doesn't have a consistent signature. Behavioral detection (monitoring for malicious actions) still works, but even AI malware might imitate normal user behavior to blend in. The "malicious GPT" concept in CrowdStrike's report notes that an altered GPT model could churn out all sorts of attack artifacts like phishing content or code.[10] It's not far-fetched to imagine a custom "cyberattack model" trained by a nation-state to assist their hackers in real operations.

In summary, AI can increase the **speed, stealth, and sophistication** of malware attacks. China's offensive cyber units could leverage these capabilities to carry out campaigns that spread faster, hit harder, and stay hidden longer in U.S. networks. It raises the stakes for defenders: we may be fighting malware that effectively *thinks* about how to outmaneuver us.

13. AI-Evasion of Security Detection and Response

On the flip side of offense, AI can help attackers thwart the defenders' AI. Many cybersecurity tools now incorporate machine learning for anomaly detection, user behavior analytics, spam filtering, etc. Chinese attackers can employ **counter-AI tactics** to evade or confuse these defenses, ensuring their intrusions remain undetected.

One method is using AI to generate inputs that specifically bypass security filters. This is similar to adversarial examples (vector 8) but focused on security systems. For instance, email filters that use ML to catch phishing might be vulnerable to AI-crafted emails that skirt those patterns. An AI could trial-and-error different phishing email variants through a spam filter (or train a surrogate model of the filter) until it finds a version that is not flagged but still carries the malicious payload. This results in **phishing that sails through filters** where normal attempts might be caught. A CNBC report noted a huge surge (over 1000%) in AI-created phishing emails after ChatGPT became available[15] – attackers are likely already doing this, sending more convincing emails that don't trigger keyword-based rules.

Network intrusion detection systems (IDS) and endpoint detection & response (EDR) systems are also increasingly AI-driven, identifying odd behaviors that indicate hacks. Attackers can use AI to learn what "normal" looks like to those systems and then mimic it. For example, if a model monitors login times and locations, an AI might plan its breaches to occur during typical working hours from IPs similar to legitimate ones, thus flying under the anomaly radar. If an AI monitors file access patterns, the malware could throttle or sequence its data access to match normal user activity patterns.

There's also the concept of **spoofing defensive AI**. An attacker might feed false data to an AI system to either trick it or overload it. Think of a scenario where a defender's AI is learning from system logs to detect hackers. A savvy attacker could generate tons of benign-but-weird log entries (noise) to hide their malicious actions in a haystack or even poison the detection model (related to vector 7 but here targeting the defender's AI). This could be part of a live attack: as they operate, they also jack up false alerts elsewhere to distract human analysts (possibly using AI to generate those false flags convincingly).

During incident response, AI might even help attackers adapt. Many advanced threats, including Chinese APTs, already do things like **log deletion and on-the-fly tool recompilation** when they sense they've been spotted.[14] In the future, an AI agent embedded in malware could monitor for signs of detection (e.g., its process being scanned or slower network responses indicating quarantine) and then automatically change tactics or self-hide better. We could face malware that, when discovered, polymorphs into a dormant state or cleans up traces faster than a human responder can react.

This evasion isn't purely theoretical. DARPA ran a challenge a few years ago on AI vs AI in cybersecurity (the "Cyber Grand Challenge"), illustrating automated attack and defense. Now, those concepts are gradually moving from research to real-world tools.

For U.S. defenders, dealing with AI-augmented attackers means our defense systems can't be static or naive. It's a bit like a game of chess where the opponent can see how our pieces move and adapt their strategy instantly – perhaps even predicting our incident response moves. The use of AI for evasion by nation-state hackers ups the need for **resilience and**

multi-layered security – so if one layer is fooled, another can catch the issue. It also means human intuition and hunting might regain importance; ironically, if both sides deploy AI, sometimes a smart human analyst might notice a subtle inconsistency that an automated system overlooking patterns might not.

In summary, China's hackers can be expected to deploy AI not just to break in, but to *stay in*. By outsmarting defensive AI and exploiting the predictability of automated defenses, they aim to maintain long-term stealthy access – which has always been a hallmark of Chinese operations (they favor quiet data exfiltration over noisy destruction). AI just helps them disappear even deeper into the digital shadows.

14. Intelligent Botnets and Autonomous Malware Agents

China has historically been linked to large botnets – networks of compromised computers or IoT devices used for coordinated attacks. With AI, botnets can become far more **autonomous and effective**. Envision a botnet of millions of devices (could be routers, security cameras, smart appliances) where the command-and-control is partially handed over to an AI system that dynamically assigns tasks, adapts to defenses, and optimizes the botnet's impact.

For example, in denial-of-service (DDoS) attacks, AI could help modulate the traffic from a botnet to avoid simple detection. Instead of all bots blasting at once (which is easier to detect and mitigate by traffic signatures), an AI might instruct different waves, patterns, or even content that makes the traffic blend in with legitimate usage until the critical moment. It could also target specific **weak points** of a network by analyzing how the target responds to initial probes. A human botnet herder might try a couple of methods; an AI can try thousands of micro-variations to see what yields the most damage.

Chinese groups like **Flax Typhoon** have been observed leveraging IoT devices to build botnets for persistence and attack launchpads.[1] The natural evolution is to incorporate AI to manage these vast armies. Imagine an AI-driven botnet that's not only for DDoS but can perform distributed tasks like **distributed intrusion**: different bots trying different exploits on different targets, sharing successful tactics among each other via the C2 AI brain. It becomes a self-spreading, self-teaching organism to some extent.

Another nightmare scenario: **swarm attacks**. In a kinetic conflict, China could unleash swarms of small drones or autonomous programs to overwhelm U.S. systems. AI coordination would allow these swarms (digital or physical) to coordinate without needing explicit human micromanagement. They could communicate and react to changes on the fly. If one bot finds a crack in a network, nearby bots (in network space, not geography) could pivot to exploit it, like ants finding a gap under a door.

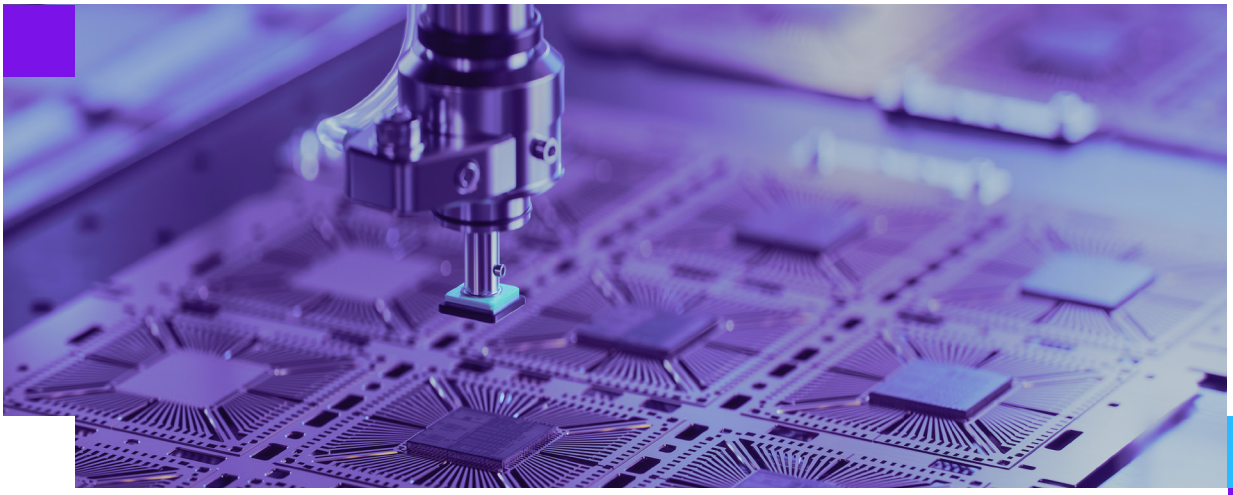
Autonomous agents could also **maintain access**. Volt Typhoon, for example, methodically re-validates its access over the years, re-compromising the same targets to ensure continued presence.[12] An AI agent implanted in a network could do this job: if an admin changes a password or updates a system, the agent can seek a new vulnerability or credential to regain what was lost, *without waiting for orders*. It's like having a persistent burglar in your house who, if you change the locks, immediately finds the spare key you forgot about.

While science-fiction sounding, elements of this are seen in malware like **botnet worms** (e.g., Conficker or Mirai) that propagate widely and adapt to some extent. The difference is the potential sophistication and distributed decision-making. Intelligence can be centralized (one big AI controlling the bots) or decentralized (each bot has enough smarts to act semi-independently). Either way, the agility and resilience of the botnet increase.

The risk to U.S. infrastructure here is multifaceted. Botnets can not only disrupt services but also serve as **cover for more targeted intrusions**. Amid a massive botnet attack, a few bots might quietly steal data or insert logic bombs while defenders are distracted. This aligns with combined operations where a noisy DDoS might mask a subtle data breach.

Countering AI-driven botnets might require AI-driven defenses that analyze patterns faster than humans can. But it also may push responses toward more drastic measures, like preemptively disabling vulnerable IoT devices or aggressive takedowns of command servers. Internationally, if a botnet of refrigerators and DVRs worldwide starts behaving like Skynet under Chinese control, we're in uncharted territory regarding rules of engagement.

In short, China's extensive access to global device manufacturing (a lot of IoT devices are made in or have components from China) combined with AI could yield **autonomous cyber weapons** that operate at scale. The U.S. must prepare for cyber incidents where the "attacker" is not a person at a keyboard, but a self-directed malicious AI coordinating thousands or millions of endpoints.



15. Compromising AI Hardware and Chips

While software and data are one side of AI, the **hardware** that runs AI – chips, GPUs, specialized accelerators – is another potential attack surface. China is both a major consumer and producer of advanced electronics. There is concern that Chinese-designed or manufactured hardware used in American AI systems could come with backdoors or hidden vulnerabilities at the **firmware or silicon level**. If so, Beijing could exploit these to undermine U.S. AI infrastructure or steal data processed by these chips.

One angle is the supply chain of AI chips. Many AI workloads run on GPUs (graphics processing units) that are predominantly made by U.S. companies like NVIDIA but often fabricated overseas (though not usually in China for cutting-edge nodes, due to sanctions). However, China is ramping up its own AI chip industry (e.g., companies like Huawei's HiSilicon, Cambricon, etc.). If Chinese AI chips make their way into U.S. products or data centers (legally or via the gray market), they could be Trojan horses. A chip could be designed to perform normally fine but have a hidden mode triggered by a specific sequence of instructions or signals that opens a backdoor. This is analogous to how some past network devices from Chinese vendors have been suspected to contain hidden admin accounts.

Even without an explicit malicious backdoor, subtle weaknesses could be introduced. For instance, a chip's random number generator (critical for cryptography) could be biased in a way known to the adversary, allowing them to defeat encryption. Or a chip might leak data through electromagnetic side channels more than typical, aiding remote eavesdropping if one knows how to exploit it. These hardware-level issues are extremely hard to detect without exhaustive testing and reverse engineering.

Another aspect is **firmware** – the low-level software that runs on hardware (like BIOS, controller firmware, etc.). Chinese companies produce a lot of components (motherboards,

disk controllers, cameras) that run firmware. Compromised firmware can subvert a system from the moment it boots, regardless of what OS is on top. Notably, the espionage group **LoJax** (believed Russian) showed it's possible to implant persistent malware in UEFI firmware. Chinese actors could do similarly or simply manufacture components with tampered firmware out-of-the-box. Reports in the past (some contested) have alleged that Supermicro motherboards made in China were tampered with to include spying chips[1] – a claim involving Chinese intelligence inserting tiny chips the size of a pencil tip on server boards (Bloomberg's controversial "*The Big Hack*" story). While that specific claim remains debated, it highlighted the feasibility of hardware supply chain tampering.

If U.S. AI researchers or companies unwittingly use compromised hardware, China could potentially **extract model data or influence computations**. Consider an AI training cluster: if some GPUs there have backdoors, an attacker might remotely collect whatever data is fed through them (like training datasets or model parameters). If it's a military AI system, that could leak highly sensitive info. Or they might inject slight calculation errors at critical moments – imagine a military simulation AI that suddenly gives a subtly wrong output due to a hardware perturbation, affecting strategic decisions.

Finally, there is the scenario of **sabotage**: in a crisis, if the U.S. relies on Chinese-made AI hardware, could China remotely disable or degrade it via built-in kill switches? It's a far-fetched but not impossible threat. At a minimum, over-reliance on foreign hardware is seen as a strategic risk, which is why there's a push for domestic semiconductor manufacturing.

The U.S. government has already banned or heavily scrutinized Chinese hardware in telecommunications (Huawei, ZTE bans 5G equipment) and in surveillance cameras (e.g., Hikvision, Dahua), citing spying fears. For AI, the issue is slightly different because the leading AI chips are not currently from China – in fact, China is trying to *import* A100 and H100 GPUs, not export them. However, as they develop their own, or for more commodity hardware, vigilance is needed. Even something as simple as a Chinese-made USB drive can have firmware that is malicious.

In summary, while software vectors abound, we shouldn't ignore the **hardware layer**. If China can't beat U.S. AI in pure tech competition, it might seek to **undermine its foundation** by compromising the building blocks. Ensuring a secure hardware supply chain and diversifying away from untrusted suppliers is thus part of the defense against AI-era cyber threats.

16. AI-Fueled Insider Recruitment and Social Engineering

Not all compromises come through digital exploits; sometimes, the oldest trick – turning an insider – is the most effective. Here, AI can assist China in **identifying, grooming, and even directly engaging potential insider threats** within U.S. organizations. By mining data (as discussed in vector 10) for personality traits and grievances, AI could flag individuals

who might be susceptible to recruitment or coercion. This augments human intelligence operations, making them more targeted.

For instance, an AI system might analyze a federal department's employee satisfaction survey (if leaked or scraped from internal systems), combine it with those employees' LinkedIn and Glassdoor posts, and find who is disgruntled or who has financial hardships (maybe visible through court records, etc.). Those individuals could be prime targets for a spy handler. In the past, Chinese espionage has recruited everyone from FBI agents to defense contractors by exploiting vulnerabilities like debts or ideological sympathy. AI can make the process of finding the "right" person much more efficient.

Moreover, AI chatbots masquerading as friendly contacts could maintain communication with a target over time – essentially, an AI-driven agent that builds rapport. Suppose a U.S. researcher is active on an international scientific forum. An AI persona (backed by real data of a fake identity) could converse with them for months, discuss research, commiserate about funding struggles, etc. When trust is established, it might suggest a lucrative opportunity – perhaps unwittingly funneling sensitive info or even inviting them to collaborate on a project that ultimately exfiltrates IP. All this could be orchestrated with minimal actual human time on the Chinese side until the hook is set.

China has a massive talent recruitment program (the Thousand Talents Plan and similar), which sometimes blurs the line between legitimate academic collaboration and IP theft. AI can refine these efforts by evaluating which scientists or experts are most likely to jump ship or share data and how to entice them effectively (tailoring pitches to their interests, as gleaned from social media or publications).

We could also see AI being used to **forge or handle initial outreach** en masse. Sending spearphishing emails to employees with access, inviting them to a conference in Beijing with all expenses paid, for example. The content of those invites and the subsequent interactions could be AI-personalized to maximize the response rate.

Another aspect is **insider threat detection**: ironically, while we worry about insiders in our organizations, an adversary might use AI to detect moles in their own ranks or as double-checks on recruited assets. But focusing on compromise: an insider doesn't have to be knowingly spying; they could be manipulated to leak info via an AI agent that pretends to be someone else (like a higher-up requesting a document, which is a mix of phishing and social engineering).

In sum, AI helps scale the human-centric side of espionage. It's not all about zeros and ones in code; AI can play psychologist, friend, and influencer. And unlike a human operative, an AI can handle thousands of potential targets in parallel, patiently waiting for one to bite. The cliché is that "people are the weakest link" in security; AI will enable adversaries

to systematically exploit that weakness on a grand scale. The U.S. counterintelligence community will need to recognize subtle AI-driven social engineering attempts and better inoculate the workforce against them.

17. Undermining Trust in AI Systems and Data

As the U.S. increasingly uses AI for decision support (from military targeting to medical diagnoses to financial modeling), an adversary can seek to **sabotage the trustworthiness** of those AI outputs. Even if they can't directly control a U.S. AI system, they can attempt to make it unreliable or to convince users it's unreliable. This can be done through a combination of some vectors we've discussed (poisoning, adversarial input, misinformation), culminating in the effect that U.S. personnel *lose trust* in their own tools or data.

One approach: cause an AI system to produce a very visible failure at a critical moment, undermining confidence in it. For instance, if Chinese hackers know an AI is used in processing intelligence reports, they might insert an adversarial example that causes it to misidentify a civilian aircraft as a hostile missile on radar. If that false alert goes up the chain and is later proven false, commanders might start doubting the AI's reliability – a doubt that China could exploit by then actually sending a hostile missile, hoping it gets dismissed as another glitch.

Alternatively, they might subtly bias data (poisoning), so the AI gives skewed advice that leads to policy or investment mistakes, which only become clear later. By then, the damage (wasted resources, missed opportunities) is done, and decision-makers might either overcorrect or abandon a useful system entirely.

There's also a psychological operation element: spreading **fear, uncertainty, and doubt (FUD)** about AI systems. China could amplify any incident of AI failure in the U.S. (through propaganda channels or fake personas) to make Americans distrust their technology. For example, if a self-driving car malfunctions, propaganda bots might push narratives about how American AI is fundamentally unsafe or easily hacked (even if they were the ones who hacked it!). The goal is to slow U.S. adoption of AI or cause internal conflicts about its use.

This vector is a bit meta – it's using all other methods not necessarily to attack directly but to create systemic doubt. In warfare, if your enemy doesn't trust their own instruments, they hesitate and make mistakes.

A historical analogy: in WW2, Allies fed false data to Nazi intelligence to undermine their trust in sources (Operation Mincemeat, etc.). Here, feeding false data to AI or making AI seem compromised could similarly mislead or paralyze.

Hugging Face dependency attacks (vector 6) could also be used to sabotage models in a

way that's later detected, leading to recalls of AI systems. If it came out that a widely used AI model had Chinese malware (like those 100 models found with malicious code[8], many organizations might pull back on using community models, slowing innovation or causing expensive rebuilds.

In the commercial space, China could sabotage U.S. AI-driven businesses to reduce their competitive edge. If an American trading firm's AI is manipulated to make bad stock trades, it not only causes financial loss but might push them to abandon AI-driven strategies that were actually beneficial, ceding ground to Chinese firms.

Finally, consider critical infrastructure: if an AI controlling part of the power grid is tricked into an error causing a blackout, regulators might ban AI automation in grids, which could prevent efficiency improvements the U.S. might need for future load management – a long-term strategic handicap.

This attack vector essentially is about **attacking the relationship between humans and AI** in the U.S. It's a subtle form of compromise – not necessarily stealing data or shutting down systems (though those can be means to the end) – but compromising confidence and consistency. The U.S. needs robust validation and fail-safes for AI recommendations and a culture of understanding AI limitations so that neither blind trust nor total rejection predominates, both of which can be exploited by an adversary.

Summary

Summary of the different AI-Driven Cyber Threats with suggested mitigation strategies:

#	Threat Area	Threat Description	Mitigation Strategies
1	Malicious Code Generation via AI Tools	AI-generated code could introduce vulnerabilities, backdoors, or insecure dependencies.	Conduct rigorous AI code audits, enforce secure coding practices, and verify AI-generated code before deployment. Use best practice software assurance.
2	AI-Powered Misinformation Campaigns	AI-generated deepfake videos, fake news, and bot-driven propaganda spread disinformation.	AI-driven fact-checking, digital literacy programs, and real-time monitoring of disinformation campaigns. Block where possible misinformation sites.
3	Data Mining of U.S. Users' Queries and Habits	Chinese AI platforms may track user searches, chat logs, and behavioral patterns for intelligence gathering.	Block untrusted AI platforms, use zero-trust architecture, encrypt sensitive queries, and monitor outbound data.
4	AI-Enhanced Spearphishing & Social Engineering	AI-crafted phishing emails, messages, and calls impersonate trusted sources with high precision.	AI-driven phishing detection, employee training, and multifactor authentication (MFA).
5	Backdoored AI Models (Trojan AI)	AI models are secretly programmed with hidden triggers to misbehave under certain conditions.	Vet and audit third-party AI models, implement secure model training, and monitor AI behavior.
6	Exploiting Trust in AI Platforms and Libraries	Attackers upload malicious AI models or poisoned libraries to public repositories.	Validate AI packages, scan repositories, and deploy software supply chain security controls.
7	Compromising AI Training Data (Poisoning the Well)	Adversaries manipulate training data to skew AI model behavior.	Use diversified, authenticated training datasets, employ data integrity checks, and secure data labeling.
8	Adversarial Attacks on AI Systems	Carefully crafted inputs deceive AI models into making incorrect predictions.	Train AI models with adversarial resistance techniques and conduct red team testing.

#	Threat Area	Threat Description	Mitigation Strategies
9	Automated Vulnerability Discovery and Exploitation	AI scans for vulnerabilities faster than human researchers, allowing mass exploitation.	Use state of the art vulnerability intelligence. Automated vulnerability management, AI-driven threat intelligence, and rapid patching strategies.
10	AI-Driven Reconnaissance & Target Selection	AI analyzes vast datasets to identify and prioritize espionage targets.	Limit OSINT exposure, use deception techniques (honeytokens), and enforce data-sharing restrictions.
11	Deepfake Impersonation & Audio/Video Spoofing	AI-generated deepfake voices and videos impersonate executives and government officials.	Deepfake detection tools, biometric verification, and secondary validation protocols.
12	AI-Generated Malware & Polymorphic Attacks	AI autonomously generates and adapts malware to evade detection.	AI-driven endpoint detection, behavioral analysis, and malware sandboxing.
13	AI-Evasion of Security Detection & Response	AI modifies attack patterns in real-time to bypass security tools.	Implement AI-driven security defenses that continuously learn and adapt, and apply behavior-based monitoring.
14	Intelligent Botnets & Autonomous Malware Agents	AI-controlled botnets self-learn and optimize attack strategies.	Traffic analysis, anomaly detection, and IoT/OT security controls.
15	Compromising AI Hardware & Chips	Backdoors in AI hardware or Chinese-manufactured chips could allow remote exploitation.	Secure supply chains, conduct hardware audits, and enforce firmware integrity checks.
16	AI-Fueled Insider Recruitment & Social Engineering	AI pinpoints vulnerable individuals and automates espionage recruitment efforts.	Insider threat monitoring, behavioral analytics, and compartmentalized access controls.
17	Undermining Trust in AI Systems and Data	Adversaries create doubt in AI decisions through sabotage, misinformation, or manipulation.	Secure AI validation, redundancy in decision-making processes, and AI audit trails.

How Armis Can Help Mitigate These Threats

Armis Centrix™, the Armis Cyber Exposure Management & Security Platform, is powered by the Armis AI-driven Asset Intelligence Engine, which sees, protects and manages billions of assets around the world in real time.

Armis' AI-driven asset intelligence engine that monitors billions of assets world-wide in order to identify cyber risk patterns and behaviors. It delivers unique, actionable cyber intelligence to detect and address real-time threats across the entire attack surface:

1. Detecting & Preventing Unauthorized Communications

- Identify and block unauthorized outbound communications to Chinese-controlled AI services (e.g., DeepSeek, suspicious LLM platforms).
- Detect and alert on devices or communications with high-risk domains.
- Use policy-based enforcement to stop AI-powered unauthorized data exfiltration.
- Use Armis Centrix™ for Early Warning and Continuous Hunt to mitigate threats early.

2. AI-Powered Anomaly Detection

- Identify unusual device behavior, lateral movement, and unauthorized access indicative of AI-driven cyber threats.
- Detects irregular network activity and deviations in device communication patterns and unauthorized device behaviors.
- Correlate AI-driven phishing or reconnaissance activities with the Continuous Hunt team.

3. Malicious Code Generation via AI Tools

- Use Armis Centrix™ for VIPR - Prioritization and Remediation to discover, deduplicate, contextualize, prioritize, assign and mitigate vulnerabilities and other security findings that may indicate or allow for malicious code development and/or execution.

4. Detecting & Preventing Unauthorized Communications

- Ensure industrial control systems (ICS)/OT systems and IoT assets are protected from AI-powered cyber attacks.

- Detect and prevent adversarial AI poisoning within firmware and connected devices.
- Monitor AI-generated malware attempting to propagate across and critical infrastructure networks.

5. Automated Vulnerability Discovery and Exploitation

- Armis Centrix™ for Early Warning leverages deception technology and can proactively get ahead of the threat and preempt an attack. The Intelligence has been proven to be ahead of industry benchmarks by 3 to 6 months.

6. AI-Powered Malware Detection & Threat Intelligence

- Detect AI-generated malware mutations and self-adapting threats by leveraging behavioral threat intelligence.
- Identify and isolate malicious AI models or backdoored software components in an organization's digital supply chain.
- Prevent AI-based adversaries from modifying detection algorithms to evade security tools.

7. Supply Chain Security & Hardware Integrity

- Ensure hardware integrity by monitoring the security of connected AI hardware, GPUs, and embedded systems.
- Detect anomalous behaviors in AI-powered hardware acceleration components for unauthorized data manipulation.
- Identify firmware-based AI attacks and unauthorized firmware modifications before exploitation.

8. Continuous Threat Intelligence & Monitoring

- Monitor AI-powered reconnaissance and OSINT threats.
- Proactively track and block compromised endpoints or AI-generated adversarial traffic in real-time.
- Correlate AI-powered cyberattacks with geopolitical intelligence to anticipate emerging threats.

Conclusion: A New Era of Cyber Threats

China's interest in AI as a strategic asset is clear. Beijing's leaders have openly stated their goal to lead in AI technology by 2030, and they view dominance in this field as key to economic and military power. The Chinese Communist Party's blend of state-driven tech innovation and espionage – often termed the **"Digital Silk Road"** when abroad and **"military-civil fusion"** at home – means advances in AI can quickly translate into new tools for intelligence and conflict. As we've explored, these tools span a remarkable array of attack vectors, from **malicious code and model tampering** to **deepfake deception and automated hacking swarms**. The playing field of cyber operations is poised to become more automated, more scalable, and, in some ways, more insidious as human hackers give way to AI-augmented ones.

Crucially, many of the threats discussed are not theoretical musings about a distant future. We already see the early signs of AI being weaponized:

- Chinese influence networks are using **AI-generated avatars** and images to meddle in U.S. discourse.[4]
- Security analysts found **backdoored AI models** lurking in public repositories.[8]
- Research proves AI can craft phishing lures that fool a majority of people.[6]
- Chinese state hackers like Volt Typhoon and Salt Typhoon have demonstrated the patience, stealth, and access to carry out complex attacks – AI will only amplify such capabilities.[1]
- Even top officials like FBI's Chris Wray are publicly warning that China could **"use AI to find and exploit vulnerabilities and conduct more sophisticated spear-phishing"** to further its espionage.[13]

The convergence of China's formidable cyber army with cutting-edge AI means the U.S. must prepare for a new kind of confrontation. This isn't a sci-fi scenario of robots waging war, but rather AI quietly integrated into the tactics of spies, hackers, and propagandists.

So, how can the U.S. mitigate these threats? A few key steps emerge:

- **Bolster AI Security:** Investments in AI safety research, model auditing, and secure AI development practices are needed. Just as we hardened software development with code reviews and DevSecOps, we need analogous practices for training data vetting, model verification, and supply chain checks for AI. For instance, scanning AI models for hidden triggers or malware (as JFrog did[8]) should become routine.

- **Promote Trusted AI Platforms:** Reducing reliance on foreign or un-vetted AI services is critical. If "DeepSeek" ever materialized as a real product, perhaps the U.S. would treat it like Huawei – with extreme caution or bans for government use. Developing and promoting trustworthy domestic or allied alternatives (and standards for transparency) can mitigate data siphoning risks.
- **AI-Augmented Defense:** Just as attackers use AI, defenders can too. AI can help monitor networks, detect deepfakes, and predict an adversary's next moves by analyzing patterns. Organizations should incorporate AI-driven security tools but remain aware of their limits (to avoid the evasion pitfall). A combination of human expertise and AI analysis – centaur teams, so to speak – might be the winning formula.
- **Public Awareness and Resilience:** The general public and workforce should be educated about AI-enabled deception. Phishing training now should include deepfake awareness (e.g., verify unusual requests via secondary channels). Media literacy programs can help inoculate against AI-generated propaganda. Essentially, boosts the "immune system" of society to recognize when something's off, even if it looks or sounds plausible.
- **International Norms and Alliances:** Addressing nation-state AI threats will also require diplomacy. The U.S. and allies (Five Eyes and beyond) are actively sharing intel on Chinese cyber operations[13] – this must extend into AI threats. Perhaps norms can be established, at least among responsible states, against certain high-risk actions (like tampering with AI in critical infrastructure). However, expecting adversaries to abide is wishful; nonetheless, coalition defense (e.g., joint takedowns of botnets, collective sanctions for cyberattacks) can impose costs on perpetrators

As we stand on the cusp of this AI-driven security paradigm, clarity and vigilance are paramount. Each of the vectors detailed above could fill its own playbook of defensive measures. But the common thread is **maintaining trust** – trust in our systems, in our information, and in our ability to control the technology we create. China's strategy, as demonstrated historically and extrapolated with AI, often aims to quietly erode that trust and exploit the gaps.

In a sense, the U.S. must navigate a dual-use dilemma: AI holds great promise to improve lives and security, yet in the wrong hands, it becomes a force multiplier for threats. Striking the balance means encouraging innovation but with robust safeguards and a keen eye on those who would twist innovation toward nefarious ends. It's a challenging task, but not an insurmountable one.

To conclude, the same playbook that applied to earlier cyber conflicts applies here: **know your adversary, harden your defenses, educate your people, and stay a step ahead in technology**. The difference now is the speed and scale at which AI enables the adversary to act. This in-depth look at China's potential AI-enabled attack vectors should serve as a wake-up call. The threats are diverse, technical, and evolving – but being forewarned is forearmed. By anticipating how AI might be used against us, we can adapt and ensure that the coming AI age is defined not by new vulnerabilities but by the resilience and ingenuity with which we secure our nation against them.

About the author

Andrew Grealy, Head of Armis Labs

Andrew Grealy is Head of Armis Labs, the dedicated research practice at Armis. Equipped with state-of-the-art tools and methodologies that leverage one of the largest data sets in the world, Armis Labs conducts in-depth analyses of evolving threats, both in the pre-emergence stage and "in the wild" stage of an attack. Andrew has a vast background in AI, threat detection and threat intelligence. He was most recently the CEO and Co-Founder of CTCI, which was acquired by Armis in February 2024. Andrew is also an advisor to multiple AI and cybersecurity companies.

Sources

1. Paul Asadoorian, *"The Rise of Chinese APT Campaigns: Volt Typhoon, Salt Typhoon, Flax Typhoon, and Velvet Ant,"* Eclypsium, Oct. 24, 2024 (The Rise of Chinese APT Campaigns: Volt Typhoon, Salt Typhoon, Flax Typhoon, and Velvet Ant - Eclypsium | Supply Chain Security for the Modern Enterprise) (The Rise of Chinese APT Campaigns: Volt Typhoon, Salt Typhoon, Flax Typhoon, and Velvet Ant - Eclypsium | Supply Chain Security for the Modern Enterprise)

2. Jessica Ji et al., *"Cybersecurity Risks of AI-Generated Code,"* Center for Security and Emerging Technology, Nov. 2024 (Snyk's 2023 AI-Generated Code Security Report | Snyk) (Snyk's 2023 AI-Generated Code Security Report | Snyk)

3. Snyk Security, *"AI-Generated Code Security Report 2023,"* Feb. 2023 (Snyk's 2023 AI-Generated Code Security Report | Snyk) (Snyk's 2023 AI-Generated Code Security Report | Snyk)

4. Voice of America (Lin Yang), *"Probe finds Beijing seeking to mislead, sow distrust ahead of US election,"* Sept. 28, 2024 (Probe finds Beijing seeking to mislead, sow distrust ahead of US election) (Probe finds Beijing seeking to mislead, sow distrust ahead of US election)

5. Graphika Report, *"Deepfake It Till You Make It – Pro-Chinese Actors Promote AI-Generated Video Footage of Fictitious People,"* Feb. 7, 2023 () ()

6. Malwarebytes (Pieter Arntz), *"AI-supported spear phishing fools more than 50% of targets,"* Jan. 7, 2025 (AI-supported spear phishing fools more than 50% of targets | Malwarebytes) (AI-supported spear phishing fools more than 50% of targets | Malwarebytes)

7. U.S. Department of Homeland Security, *"Data Security Business Advisory,"* Dec. 22, 2020 (Data Security Business Advisory: Risks and Considerations for Businesses Using Data Services and Equipment from Firms Linked to the People's Republic of China) (Data Security Business Advisory: Risks and Considerations for Businesses Using Data Services and Equipment from Firms Linked to the People's Republic of China)

8. Dark Reading (Elizabeth Montalbano), *"Hugging Face AI Platform Riddled With 100 Malicious Code-Execution Models,"* Feb. 29, 2024 (Hugging Face AI Platform Riddled With 100 Malicious Code-Execution Models) (Hugging Face AI Platform Riddled With 100 Malicious Code-Execution Models)

9. CrowdStrike, *"Most Common AI-Powered Cyberattacks"* (web article), 2023 (Most Common AI-Powered Cyberattacks | CrowdStrike) (Most Common AI-Powered Cyberattacks | CrowdStrike)

10. CrowdStrike, *"Adversarial Machine Learning"* (excerpt), 2023 (Most Common AI-Powered Cyberattacks | CrowdStrike) (Most Common AI-Powered Cyberattacks | CrowdStrike)

11. Trend Micro, *"Unusual CEO Fraud via Deepfake Audio Steals $243,000,"* Sep. 5, 2019 (Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company | Trend Micro (US))

12. Federal Communications Commission, *"Impacts of Salt Typhoon Attack and FCC Response"* (Fact Sheet), Dec. 2024 (The Rise of Chinese APT Campaigns: Volt Typhoon, Salt Typhoon, Flax Typhoon, and Velvet Ant - Eclypsium | Supply Chain Security for the Modern Enterprise)

13. SC Media (Simon Hendery), *"FBI boss slams 'unprecedented' Chinese cyberespionage and IP theft,"* Oct. 18, 2023 (FBI boss slams 'unprecedented' Chinese cyberespionage and IP theft | SC Media) (FBI boss slams 'unprecedented' Chinese cyberespionage and IP theft | SC Media)

14. CISA Alert AA24-038A, *"PRC State-Sponsored Actors Compromise U.S. Critical Infrastructure (Volt Typhoon),"* Feb. 2024 (PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure | CISA) (PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical Infrastructure | CISA)

15. CNBC, *"ChatGPT creating huge increase in malicious phishing email,"* June 2023 (AI like ChatGPT is creating huge increase in malicious phishing email) (Stat on 1265% increase in AI phishing emails).

16. Cobalt, "Backdoor Attacks on AI Models" Dec. 20, 2023 (Backdoor Attacks on AI Models).

17. Qwiet, "HuggingFace Exploit Highlights AI Supply Chain Vulnerability" Feb. 28, 2024 (HuggingFace Exploit Highlights AI Supply Chain Vulnerability).

18. NSFOCUS, "AI Supply Chain Security: Hugging Face Malicious ML Models" March 5, 2024 (AI Supply Chain Security: Hugging Face Malicious ML Models).

19. PROTECT AI, "Supporting the safe and secure usage of the world's largest AI/ML Model Repository" Oct. 25, 2024 (Supporting the safe and secure usage of the world's largest AI/ML Model Repository).

20. Forbes, "Google Claims World First As AI Finds 0-Day Security Vulnerability" Nov. 5, 2024 (Google Claims World First As AI Finds 0-Day Security Vulnerability).

21. NCSC, "The near-term impact of AI on the cyber threat" Jan 24, 2024 (The near-term impact of AI on the cyber threat).

22. LastPass Labs, "Attempted Audio Deepfake Call Targets LastPass Employee" April 10, 2024 (Attempted Audio Deepfake Call Targets LastPass Employee).

ARMIS LABS

About Armis Labs

Armis Labs, a division of Armis, is a team of seasoned security professionals dedicated to staying ahead of the ever-evolving cybersecurity landscape. With a deep understanding of emerging threats and cutting-edge methodologies, Armis Labs empowers organizations with unparalleled visibility and expertise to protect against the threats that matter most, right now.

At the heart of Armis Labs lies a formidable research powerhouse, where experts investigate the latest trends and tactics employed by cyber adversaries. Armed with access to over 5 billion profiled assets and state-of-the-art tools and methodologies, the team at Armis Labs conducts in-depth analyses of evolving threats both in the pre-emergence stage and "in the wild" stage of an attack.

+5 Billion

Core to Armis Labs is our Asset Intelligence Engine. It is a giant, crowdsourced, cloud-based knowledge base - the largest in the world, tracking over five billion assets - and growing. It powers Armis Labs with unique, actionable cyber intelligence to detect and address real-time threats across the entire attack surface.

Armis Labs security practitioners are utilizing cutting edge technology that include dynamic honeypots, incident forensics, reverse engineering, dark web monitoring, and human intelligence to proactively identify and mitigate threats before they manifest. Leveraging advanced AI/ML technologies, Armis Labs' proactive threat detection capabilities enable organizations to stay one step ahead of cyber adversaries, minimizing the risk of potential breaches while stopping potential damage before it occurs.

Armis Labs is dedicated to providing organizations with the tools and expertise they need to defend against the threats that matter most, right now. With comprehensive threat intelligence, proactive threat detection capabilities, and seamless integration into existing security workflows, Armis Labs empowers organizations to stay ahead of cyber adversaries and protect their most critical assets.

Armis, the cyber exposure management & security company, protects the entire attack surface and manages an organization's cyber risk exposure in real time.

In a rapidly evolving, perimeter-less world, Armis ensures that organizations continuously see, protect and manage all critical assets - from the ground to the cloud. Armis secures Fortune 100, 200 and 500 companies as well as national governments, state and local entities to help keep critical infrastructure, economies and society stay safe and secure 24/7.

Armis is a privately held company headquartered in California.

1.888.452.4011

Website

Platform
Industries
Solutions
Resources
Blog

Try Armis

Demo
Free Trial