6 War, Cyberwar and Hacker-AI

“War is Politics with Other Means”

War is a violent, armed conflict between nations or groups within nations. It is characterized by the use of military force, deployment of troops, and execution of tactics or strategies to defeat the enemy at almost any cost. Even post World War II War, and after the end of the Cold War, war is still being fought to achieve political or economic objectives. It asserts dominance over regions, resources, or populations despite its inevitable consequences in loss of life, destruction of infrastructure, and economic disruption.

Throughout history and even in 2023, wars have shaped events and borders. The outcome of wars is that territories were occupied, national borders were redrawn, and political or economic systems were changed. We need to acknowledge that war has significantly impacted the spread of technologies and ideas, including civil rights and even the role of women in our societies. In the past, war was often a catalyst for change, leading to significant political, economic, and social reforms. Today war of aggression is forbidden. But war itself is not completely outlawed by international law. Instead, we have several international treaties regulating the conduct of war; they aim to minimize its negative consequences, such as the United Nations Charter, the Geneva Conventions, the Hague Conventions, and the Rome Statute of the International Criminal Court.

Under any of the many definitions of war under international law, like using force or violence between states or between states and non-state actors, or if we consider the intensity of conflicts like a large number of soldiers and weapons involved, a cyberwar is usually not a war under these definitions.

Even other criteria, like the presence of hostilities, such as fighting, bombings, or the use of military force, rather than diplomatic or economic measures to resolve disputes or achieve objectives, would not immediately make cyberattacks an act of war.

A cyberwar could potentially be a war when cyberattacks disrupt or destroy critical infrastructure or military systems or cause widespread harm or damage to a country and its citizens. A high level of suffering among people, including the loss of life, and a significant impact on people’s daily life from disrupted economic, social, and political systems affecting many people are required to consider cyberwar actions as an act of war. However, there is still an ongoing debate among scholars and policymakers about how war needs to be defined.

Still, cyberwarfare and economic warfare are less destructive than traditional warfare and should therefore be treated differently. However, if power, communication, or water supply is suddenly switched off for several days, it could have irreversible, disastrous consequences. The sudden loss of essential services would disrupt daily life and cause widespread chaos and confusion; cities would become quickly inhabitable, leading to a total breakdown of society and violence among people fighting for survival.

History has taught us that all wars were costly in terms of resources, losses, and effort. These costs have a major deterring effect on all parties involved in armed conflicts. Modern wars require significant military equipment, personnel, and resources, which can be expensive to acquire, maintain and operate. Its negative economic impacts disrupt trade, damage infrastructure, and reduce investments. Social and political instabilities from war have long-term consequences for countries and communities. Empires were often crumbling under the follow-up costs that war caused. The negative uncertainty from wars deters modern societies from waging wars.

This section's title, "War is politics by other means", is a quote from a Prussian military strategist Carl von Clausewitz. He argued that war should be seen as a continuation of political activities. According to von Clausewitz, the ultimate goal of war is to achieve political objectives, and military force is simply a means to that end. This idea is still influential in some governments and is used to justify wars when there would be no serious cost to deter leaders from waging war. It seems that war, i.e., violence below a certain threshold, in pursuit of political goals is acceptable if there are no consequences. Currently, the world community uses cyber and economic warfare to increase the aggressor's cost of waging war.

We must be prepared that war will remain a factor in our world. Our defensive capabilities and deterrence preparedness are essential to prevent aggressors from using violence to pursue political goals. Yuval Noah Harari wrote: "What was normal in thousands of years of imperial history is causing outrage today. Even considering civil wars, insurgencies, and terrorism, wars have killed far fewer people in recent decades than suicide, traffic accidents, or obesity-related diseases." Wars should be made less likely to wage - technically - not just with our public outrage.

Waging war is very costly - in human lives, economically, socially, and environmentally. These costs are keeping most political leaders worrying about starting a war. But in the last decades, (kinetic) weapons got surgical and limited in their damage-related applications. Additionally, with drones, politicians saw that they could do military actions that did not risk their forces' lives. Unfortunately, technology has made some forms of war more likely to be waged.

Cyberwar 1.0

I mentioned that cyberwar does not necessarily fit the definition of war. But cyberattacks within a cyberwar could be considered an act of war if the results and damages are significant enough.

What is a cyberwar? In short, it is a conflict in which countries or groups use cyberattacks to disrupt or damage each other's computer systems and networks. In cyberwar, cyberattacks and digital sabotage are means of political coercion to instigate change. Computer and their networks are disrupted, disabled, or damaged within the adversary's infrastructure, communications, or operations. Cyberwarfare could use malware, denial of service attacks, and hacking to disrupt or gain access to sensitive information or systems.

The term "cyberwarfare" was first coined in the late 1980s and early 1990s when computers and the Internet got a more significant role in military and security affairs. I am not aware of who specifically invented the term. Still, it seems the cyberwarfare concept emerged gradually as experts, policymakers, and researchers began considering how computers could be used as weapons.

In these early days, cyberattacks were relatively simple and unsophisticated, often carried out by individual hackers or small groups. One of the first major cyberattacks occurred in 1988 with the release of the Morris worm, which infected thousands of computers and caused significant disruption.

Probably around 2007, when Russia targeted Estonia in a major cyberattack disrupting online services, causing widespread outages, digital technology and governments' and organizations' increased reliance on computer systems were seen as major vulnerabilities. At the same time, China attacked the Pentagon, which then led in 2009 to the forming of the US Cyber-Command (USCYBERCOM). NATO established its Cyber Defense Centre of Excellence (CCDCOE) in Tallinn, Estonia, in 2008. The CCDCOE is a NATO-accredited international military organization focusing on research, training, and exercises related to cyber defense. Still, it is not a US-style cyber-command but a hub for sharing knowledge and expertise on cyber defense among NATO member states and partners.

In recent years, we have seen several high-profile cyberattacks, including the Sony Pictures hack in 2014 and the NotPetya attack in 2017, which caused billions of dollars in damage. These attacks have demonstrated the potential for cyberattacks to cause widespread disruption and have raised concerns about the vulnerability of critical infrastructure. But these attacks have also shown that it is often difficult to distinguish between cyberwarfare and cybercrime using widely available spyware and ransomware. The boundaries between cyberwar and crime are often blurred. In May 2021, the Colonial Pipeline was the victim of a ransomware attack by most likely Russian cybercriminals resulting in its multi-day shutdown with consequences for critical infrastructure on the US East Coast.

There are currently three main types of cyberwarfare: propaganda/disinformation, service disruption, and espionage. They are pursued with different tactics or techniques depending on the attacker's goals or capabilities.

Social media is weaponized in different forms for distributing fake news or for creating dissent and hate within adversaries societies. Propaganda and disinformation often follow the playbook of psychological operations, also called PsyOps or psychological warfare, using emotions, attitudes, and behavior of people, groups, or entire populations to manipulate public opinion, shape perceptions, or undermine/influence an adversary's decision-making.

Additionally, malware like viruses, worms, and trojans are used to infect/damage systems, steal sensitive information, or damage a computer's normal operations. WannaCry ransomware infected 2017 computers in more than 150 countries, causing significant damage and disruption. It is believed that hackers working for the North Korean government are behind that attack. Even in 2010, Stuxnet was successfully used to sabotage the Iranian nuclear program.

In December 2020, the SolarWinds hack, a major cyberattack, was discovered, in which hackers believed to be working on behalf of the Russian government gained access to the computer systems of several US government agencies and numerous private companies through a supply chain attack on software made by the company SolarWinds. Hackers were tempering with software updates that were delivered to their customers, compromising soft-/hardware of thousands of computer systems with exploitable vulnerabilities.

Then there are phishing attacks that use fake emails or websites to trick individuals into revealing sensitive information or clicking on malicious links. They are used to steal passwords or log-in credentials or to install malware on a victim's computer. In 2016 occurred, a hack into the computer systems of the Democratic National Committee (DNC) and the Democratic Congressional Campaign Committee (DCCC) during the United States presidential election. It resulted in the theft of emails and other documents released publicly through WikiLeaks. This hack was a major factor in the 2016 US presidential election, in which Russia was accused of interfering.

Another cyberwar method is a denial of service (DoS) attack to overload or disable a website, company, or governmental websites by flooding network traffic and disrupting the availability of a service. They are launched from multiple devices working in concert, which makes it difficult to defend against.

Other considered forms of cyberwar are aimed at what wars usually do. They sabotage an adversary's systems or infrastructure. They use malware, hacking, or other tactics to damage or destroy equipment or data and create indiscriminate destruction for innocent casualties. They are designed to increase the cost of war.

An act-of-war-level cyberattack could target the physical infrastructure of a country, like power plants, transportation, or communication networks, and cause widespread disruption or damage to essential services. So currently, we should ask who is a possible cyberwar target. This is certainly the military, the critical infrastructure, the entire financial system, the governments, many private organizations, and individuals with knowledge or skills. This list seems to be ordered, and it is. Targets on the top of the list are more relevant for the conduct of war and the protection of many people within hostilities; earlier mentioned means more critical to be defended.

In November 2018, the Cybersecurity and Infrastructure Security Agency (CISA) for formed as a federal agency within the United States Department of Homeland Security (DHS). Its mission is to protect the nation's critical infrastructure from cyber-threats, improve the security and resilience of the cyber ecosystem, and provide guidance and assistance on cybersecurity matters. CISA is also involved in incident response and collaboration with international partners; it serves as a resource for the public, guiding how to protect against cyber-threats and what to do in the event of cyber incidents.

CISA compiled a list of critical infrastructure components essential to a society's functioning:

1. The power grid generates, transmits, and distributes electricity to homes and businesses.
2. Water systems / networks to produce or distribute and treat water for homes and businesses.
3. Transportation systems support the movement of people and goods, such as roads, highways, airports, air traffic control systems, train schedules, and shipping systems / ports.
4. Telecommunications networks support the transmission of data and communications like telephone, Internet, and satellite systems.
5. Financial systems / networks like banks, stock exchanges, and payment networks facilitate financial transactions and accurately manage people's accounts
6. Healthcare systems support hospitals, pharmacies, and medical supply chains.
7. Agricultural systems for food to producing, processing, storing, and timely distribution
8. Energy systems / networks to produce and distribute include oil and natural gas pipelines, electrical grids, and renewable energy systems.
9. Government's systems / networks to communicate, plan operations and provide services to their citizens - including recording deeds or public records of property ownership

Being on this list means that an organization or sector is considered essential to the functioning of a country and its ability to respond to cyber-threats; they are considered to have a high level of vulnerability to cyberattacks and are therefore prioritized for protection. Some organizations may receive additional support and resources from the government to help them improve their cyber defenses. The goal of governments seems to be to have certain cyber security standards and guidelines to reduce their risk of being targeted by cyberattacks. Attacks on critical infrastructure are closely monitored, but this does not mean an organization/sector is immune to cyberattacks.

When will we know that we are in a cyberwar?

There is no theory of cyberwar or an established legal framework or rules for declaring or conducting such conflict. Digital sabotages are often covert and never publicly acknowledged. Still, signs will indicate that we are under attack, like widespread disruptions in critical infrastructure (power, transportation, or communication), large financial losses through theft, fraud, or disruption of operations, and compromised data leading to multiple scandals and crises in leadership.

Currently, cyberwar does not fit into a legal framework that establishes rules for declaring or conducting a cyberwar. Some legal principles and frameworks may be relevant in cyberwarfare, like the international humanitarian law (IHL), also known as the law of armed conflict. It provides rules and principles to regulate armed conflict and protect civilians and non-combatants. The only enforceable tool in cyberspace is currently national laws that criminalize cyberattacks and give law enforcement and intelligence agencies the right and obligation to make private organizations and individuals accountable for their actions.

The players with advanced cyberwarfare capabilities are the United States, Russia, China, North Korea, Iran, and Israel:

- The United States is believed to have the world's most advanced cyberwarfare capabilities. The US military and intelligence agencies have established a cyber-command focused on cyberwarfare. The US government has invested heavily in cybersecurity research and development. They are accused of using cyberattacks for offensive purposes, including the Stuxnet attack against Iran's nuclear program.
- Russia has sophisticated cyberwarfare capability for defensive and offensive purposes. Russian hackers have been implicated in high-profile cyberattacks, including the 2016 US election interference and the NotPetya ransomware attack. The Russian government has also been accused of using cyberattacks to target political opponents and dissidents.
- China has well-developed cyberwarfare capability; they have been accused of using cyberattacks for espionage and offensive purposes. Chinese hackers have been implicated in high-profile cyberattacks, including the 2014 OPM (US Office of Personnel Management) data breach, with sensitive, personal data of millions of current and former govern-

ment employees, and the 2017 Equifax data breach. The Chinese government is also using cyberattacks to target political opponents and dissidents.

- North Korea has advanced capability to use cyberattacks for espionage, theft, and sabotage. North Korean hackers have been implicated in several high-profile cyberattacks, including the 2014 Sony Pictures hack and the 2017 WannaCry ransomware attack.
- Iran is assumed to have a growing cyberwarfare capability for defensive and offensive purposes. Iranian hackers are connected to several high-profile cyberattacks, including the 2012 attack on Saudi Arabian oil company Aramco and the 2013 attack on the Sands Casino in Las Vegas. They are using cyberattacks to target political opponents and dissidents as well.
- Israel likely has very sophisticated cyberwarfare capabilities. They are accused of using cyberattacks for defensive and offensive purposes. The most high-profile cyberattack they were accused of was the 2010 Stuxnet attack on Iran's nuclear program. Israel has private companies like the NSO-Group that developed advanced spyware for smartphones (Pegasus) for spying on political opponents and dissidents.

When these countries have cyberwar programs, why have others none or much less?

We could speculate that other countries don't have the resources or expertise as they lack significant technical resources and a sufficient pool of qualified personnel. There is also the concern of a potential backlash from other countries or international organizations. Some countries have legal or ethical constraints that prevent them from developing or using cyberweapons for offensive purposes or keeping their development secret for strategic reasons. As a result, it isn't easy to know which countries have cyberwarfare programs and which do not.

However, if countries are discussing governmental trojans in their fight against criminals, they have effective cyberwar capabilities, like technical expertise, to develop cyberweapons by exploiting vulnerabilities. Other criteria for cyberwar capabilities are the infrastructure to operationally plan, prepare and carry out attacks, including gathering intelligence about potential targets. For western countries, it is important to have the legal and policy frameworks that govern the use of cyberattacks and provide guidance on when or how such attacks can be used.

There is also often the discussion on who has an advantage in a cyberwar: the attacker or defender. The balance between them shifts over time; both have strengths and vulnerabilities. But there are a few general observations about the relative advantages/disadvantages of attackers and defenders: Attackers have the initiative in a cyberattack and an advantage in terms of what they choose as the targets or tactics. They can surprise with new techniques or exploit vulnerabilities, and they are specialized, giving them expertise and focus. Defenders have an advantage due to multiple layers of defense (firewalls, intrusion detection, and user education). If prepared, they could quickly recover (e.g., via backups) as they have resources/ infrastructure to respond/mitigate consequences from an attack.

Some final thoughts. Cyberwarfare should not just be a concern for governments and militaries. Private companies/individuals are at risk, which can lead to data breaches and financial losses. Our increasing reliance on technology, the proliferation of interconnected devices, and the development of new technologies, like AI and the Internet of Things (IoT), have driven the development and evolution of cyberwarfare. As technology advances, we will likely see even more sophisticated and complex cyberattacks in the future.

In response to the growing threat of cyberattacks, countries and organizations have developed a range of cyberdefense strategies; they are engaged in international cooperation to address the issue. Cyberattacks are a form of asymmetrical warfare. A cyberattack allows smaller/weaker states to level the playing field against larger or more powerful states. However, cyberattacks are difficult to trace to specific actors due to the use of proxies, encryption, and tactics that mask the source of an attack. It is currently the consensus that the risks of cyberwarfare cannot be eliminated; therefore, efforts were made to mitigate the risks and promote a safer and more secure online environment.

Finally, it seems that human intelligence, i.e., human spies in critical positions, still gives decisive hints on who should be blamed for a cyberattack. For an outsider, it seems that digital forensics is just being used as a convenient tool to make the accusations publicly appear more credible. Realistically, digital forensics alone is unreliable because the evidence can be easily tampered with or manipulated, making it difficult to determine its authenticity and integrity.

What is Cyberwar 2.0

Overview

This book was written primarily because of concerns about Hacker-AI and its application in cyberwar. Besides espionage, using cyber-tools, I have defined Cyberwar 1.0 actions as war scenarios about disrupting and destroying computer systems/services within critical infrastructure; more candidly, their goal is to increase the cost of war.

I have defined Cyberwar 2.0 as the active use of advanced malware by attackers that are not using tools to destroy or deactivate IT services indiscriminately. I will show how Hacker-AI could become a cyberweapon that reduces the cost and damage of war (significantly). In short: Cyberwar 2.0 is making the cost of waging war irrelevant, i.e., non-existent. The primary target of Cyberwar 2.0 is the decapitation of a country's government or civil society.

The following features will present a framework for the underlying capabilities that facilitates Cyberwar 2.0 activities. It will use the features/tools of Hacker-AI that were already discussed in the previous chapter. They are summarized under the following war-relevant capability categories:

(1) Surveillance.

Smartphones collect reliable, comprehensive intelligence on all citizens/organizations without the attacked country detecting that. After starting a covert attack on all smartphones, prioritization could quickly turn the surveillance on relevant people as defined by filter criteria. This data gathering (audio/text/location) will create an accurate, comprehensive, cross-referenced model of roles, responsibilities, and motivations of everyone relevant in the attacked society. Even people without a surveilled smartphone will be categorized. Malware could even look for resumes on user devices. But it will look for individual pressure points used for intimidation/coercion or later enforcing societal compliance using AI trained in detecting that. Surveillance uses the public power supply, telecommunication, the Internet, and many untouched freedoms (speech/protest) for attacker's advantage. As a reminder, there is already malware on smartphones (Pegasus/NSO-Group) that could make public surveillance infrastructure (CCTV) redundant. Surveillance will give a comprehensive picture of all IT devices and software from which the assailant could infer exploitable vulnerabilities using malware generated by Hacker-AI.

(2) Selective access denial (denied services for specific people).

Communication, Internet, and power supply are fully available for all, primarily for the surveillance of relevant people. Widespread outages from disrupting critical infrastructure are considered counterproductive, as this could unnecessarily change people's daily routines. Even denied access by some people does not need to be perceived as a malicious targeted attack or a demonstration of power. Instead, a denied access could come more innocently via unreliable services, temporary outages, or failures, all caused by devices' hidden software/malware features (unbeknown to users). On webpages, the presentation of text could be changed inconspicuously, which effectively disables the pages, e.g., by suppressing the separate CSS files (Cascading Style Sheets) that are used to define the look and formatting of web content or by disabling some JavaScript files required for effective user interactions.

(3) Directly intimidating people.

Civilians are used and intimidated, not indiscriminately or accidentally killed. With access to smartphones or IT devices, attackers could bypass all physical/logistical barriers in delivering personalized threats. Not even a phone call or text message from assailant's territory is required. A threatening chat (audio or text) with local AI bots could be more effective than a call from a real person. Threats provided with talking AI (bots) could give victims the impression that they have fewer chances to remain undetected or forgotten if threats are ignored. Also, tracking these victims and using their compliance data makes it more difficult for contacted citizens to resist their recruitment as spies or collaborators over time. Sooner or later, people are scared into collaboration and acceptance. No data traces or witnesses of these interactions would be left behind that could account for what happened. The threats could automatically contain and reveal knowledge gained from surveillance and drones showing that violence could likely be a consequence of non-compliance.

(4) Realtime Deep-Fakes, redefining truth.

If real-time announcements are made, only a very small group knows if recordings are genuine, modified in real-time, or completely fabricated. No news reporting can be trusted if real-time audio/video calls can be faked. Even documentaries from past events could be faked systematically; soon, AI could automatically give videos a new narrative. Disinformation does not need to be perfect to create confusion among people. Publications of laws, rules, and regulations are all made available digitally for convenience; they could be modified to serve an assailant's agenda. Who would catch this problem if it is not within the memory of people working with these rules and regulations regularly? Surveillance and result reporting could quickly show who needs another form of persuasion to comply with assailant's agenda fully or who needs to be replaced. The hiring, firing, promotion, or demotion could be done with automated (faked) messages. Callbacks could probably be handled automatically soon via deep-fakes. Organizing resistance against a coordinated cyberattack, in which we don't know what is true or what are lies, is very difficult and potentially impossible.

(5) Reduction of costly consequences of a typical war.

So far, all wars are extremely costly - in human lives, economically, socially, and environmentally. In Cyberwar 2.0, there is no reason to create damage or even destroy anything. Instead, the goal could be to prevent anything that is detrimentally costly, including sabotage and possible penalties from economic sanctions, via proactive and preventative espionage on every detected threat. In Cyberwar 2.0, every physical damage or destruction is counterproductive. Massive threats, including to the well-being of someone's family, should deter non-compliance with demands; made-up news on freak accidents/events could prove the seriousness of these threats. There is no reason to use ransomware or cyber vandalism because this could undermine other stealthy operations. With less destruction and fewer preventable costs, no costly disruption within the occupied country or unmitigated sanctions should devalue the spoil of war. Ideally, most civilians in an attacked country should not interrupt their daily routine or be made even aware that their country is in an existential Cyberwar

2.0. A change in government, bureaucracy, or security apparatus does not need to leave a dip in the domestic GDP. Additionally, with additional lead time, consequences from sanctions could be mitigated with advanced preparation.

(6) Misdirection.

Acknowledging the truth about the capabilities of Hacker-AI and its use in waging a Cyberwar 2.0 would likely lead to a global shock. Countries and their citizens would face vulnerabilities from the dependence on smartphones and other IT devices. There is the realistic prospect that other countries could be next to lose their sovereignty and freedom. Playing down the events and threats from Hacker-AI is essential in continuing our daily life. Proving who is responsible for a simple cyberattack, i.e., when the stakes are not so high, is (already) extremely difficult. Leaving breadcrumbs to implicate a less sophisticated patsy could help reduce the temperature in public outcry. Pointing to private / criminal organizations that use only a limited amount of malware to decapitate their government could be seen as an internal affair. Operators of Hacker-AI would know which data traces they must leave purposefully and intentionally behind, designed to misdirect digital forensics. "Cui Bono" (who benefits) could narrow the list of culprits. Cybercriminals are easy to blame. Misdirection could give assailants more time to undermine possible retaliation capabilities.

Hacker-AI and Cyberwar Requirements

Hacker-AI could have both centralized and decentralized features. Hacks or exploits executed on the attacked systems are probably developed in centralized computer facilities with data (i.e., apps) from bought devices, covertly uploaded from several attacked instances, or with data from tech libraries.

Malware generated by Hacker-AI could vary on occupied devices. It doesn't need sophisticated AI features on the attacked systems, only a smart / modifiable low-level code platform that is easily extendable and can hide its code executions. The deployed malware could consist of a small, adaptable core; it could be extended with code for its mission. Cyberwar operators know some details about a device from Cyber Reconnaissance before malware enters via Cyber Beachhead and then quickly hides within a Cyber Cradle. It gets updates undetected by the main OS or security tools via communication mainly used by humans, called Cyber Whispering; these ubiquitous communication methods are already encrypted for human privacy. Malware could carefully remove all data traces after it has been used it.

The occupying malware operates likely outside / below existing (OS) permissions. That means it would have three main tasks:

(i) hide activities from the main OS,

(ii) hide / change its code / configuration against advanced detection / forensics, and

(iii) receive/request covert instructions from the outside on what to do.

A fourth requirement might be to connect with neighbors in occupied networks, share data or explore unoccupied devices and conquer them.

The first two features are within Hacker-AI's DNA: never getting caught. Once a single malware instance might be caught or analyzed (i.e., revealed as something generic, adaptable) and reversed engineered, all similar, relevant (or possibly endangered) instances on all other non-probed systems with the same malware are changed to stay undetected/unassailable. Detectable patterns or reused (i.e., cloned) code would make detecting Cyber Ghosts easier. This side effect can be eliminated by creating diversity so that the damage from a few successful detections is limited. However, intentional detectability is a cyberwar tool that can be purposefully used for misdirection and plausible deniability.

The installed malware will not be autonomous or fully auto-aware of unforeseen circumstances that could reveal its existence. It will only respond to known or anticipated threats automatically; otherwise, malware is part of a much larger swarm of software instances that follow an attack plan managed by an attack (synchronization) management. Some malware instances may have standing orders and wait until they are changed, while others are used for limited missions and then retired after missions are accomplished. Therefore, the third feature would allow attackers to operate undercover while doing almost anything on the occupied devices.

The assumption that users must have done something wrong (like clicking a link, etc.) to have their system infected with malware is wrong. Vulnerabilities, like the one used by Pegasus spyware to allow click-free installation, are certainly not one-offs. Finding new vulnerabilities that could be used click-free is what Hacker-AI would be used for.

Not all systems are directly connected to the Internet. Some systems have an air gap, a physical separation between computers or networks. Air gaps create an additional layer of security by physically isolating systems. Air gaps disconnect systems from any network or external connections and prevent any data transfer or communication in or out of the system; it is not reachable by any unauthorized user or malware via a network, USB, or any other physical means. This makes their use and compromising them more difficult, but not impossible. Humans can be coerced to bypass or bridge many of these air gaps. Sooner or later, software code and hardware details of all devices or systems will be known to Hacker-AI; this will apply to weapon systems and include also their command and control.

Additionally, using older phones (with non-smartphone features), like burner phones, may reduce defenders' problems with (comprehensive) audio surveillance. It would not solve the problem with selected malware-triggered service denial or that people can be intimidated directly via phone calls. Even 20 years old mobile phones have operating systems that could be modified, which means they could also host malware that could trigger malware on smarter (potentially more isolated) devices nearby.

Cyberwar as Consequence of Hacker-AI

The most distinct difference to existing cyberwar scenarios is that Cyberwar 2.0 is an interactive data operation in which the collected information is actively used to avoid costly destruction or disruption. Enemy’s infrastructure is used to get real-time intelligence and lift the fog of war via reliable feedback data. Disabling defenders’ capabilities can be done using malware that is less obvious and more targeted. The population is intentionally kept in the dark about their country being occupied covertly while the government is being exchanged in a coup.

Contrary to the mainstream approach of conducting war, I assume that power supply, Internet, and communication are available and that this is more advantageous for attackers than defenders. The Cyberwar 2.0 concept assumes that every destruction is counterproductive for the attack and the aftermath. The goal of cyberwars is solely defined politically or economically. If eliminating an economic competitor is the goal, then destruction might be the tool of choice. Destruction seems more decisive or convenient; it is more easily detectable. But these are all short-sided considerations. If war is a battle of will and logistics, then there is nothing more effective than using malware to target enemies’ minds and logistics.

Malware can surveil phone calls and other smartphone activities, locations, or proximity or grab resumes from users’ eMail. People’s phone data will reveal a person’s role, status, motivation, and likely personal pressure points. The audio could be transcribed on devices, and surveillance data could be aggregated and compressed into small, inconspicuous data packages uploaded to 1000s of servers outside the target or assailant’s country covertly. The Hacker-AI-generated malware could easily use many existing features already installed on the user’s system.

The assailant could automatically derive or generate detailed plans with actions to manipulate institutions covertly from the uploaded surveillance data. To win a cyberwar, attackers must identify key people, incl. possible replacements, and create surveillance bubbles around them. Also, clerks or workers relevant to the execution of tasks could be surveillance targets. Total surveillance of everyone is unnecessary, but still, it could happen to everyone.

With methods of selectively denying access to information and communication, a coordinated malware attack could effectively decapitate governments and society without being noticed by outsiders. Deep-fakes can fill the gaps and contribute to confusion or misinformation if required. By dominating the information space, the assailant could establish new executive teams with people it has vetted via surveillance and/or intimidation.

Hacker-AI used in Cyberwar 2.0 must have additional software with mission-related features. Fully committed state actors could use governmental resources and their institutional experience in using laws, bureaucracies, and the security apparatus to define concrete enough operational goals for the aftermath. After government overthrow/regime change and with state-supported surveillance, Cyberwar 2.0 could quickly turn into a police operation to fortify gains. In the long term, surveillance tools like the social credit scoring systems in which behavior and actions are used to determine access to services and privileges. These systems establish and reward self-censoring behavior among citizens. Additionally, in an environment under total surveillance, committing sabotage or terrorism will be more difficult.

Cyberwar 2.0 can isolate and disarm the opposing military or security (police) forces loyal to the decapitated government or society. Once most IT devices are occupied by malware from assailant's Hacker-AI, there is no hiding from covert (or overt) surveillance; arrests can happen anywhere, anytime. Security could know in seconds where any of its new citizens reside. It would likely take only hours or days at most to identify and eliminate resistance. Dictatorships have shown (e.g., via actions against the Uyghurs) how possible acts of sabotage are proactively being suppressed with reeducation camps.

In Cyberwar 2.0, the assailant communicates directly with the citizens of the attacked country, not necessarily via phone calls, but more likely with (automated) chatbots run by Hacker-AI-generated malware. These bots could intimidate people and disrupt trust in the previous government. A silenced or decapitated government is prevented from giving orders. This vacuum is effectively filled by an attacker issuing (fake) orders.

The most likely goal of Cyberwar 2.0 is a zero physical damage government overthrow or regime change.

What is Detectable in Cyberwar 2.0?

Technically unprepared defenders will not detect Cyberwar 2.0 activities. Detection happens only if the assailant intentionally steps out of the shadow, e.g., when it interacts with people, which means it has, e.g., directly threatened people via AI bots. However, leaving evidence or witnesses that this has happened could probably be avoided by the malware. Threats serving an assailant's agenda could also be delivered via deep-fakes from someone who will later deny it. On a larger scale, cyberactivities could be blamed on an internal political struggle for power, like a coup. We should assume that Hacker-AI-generated malware remains undetected during all cyberwar phases or that it uses misdirection to blame others.

Off course, it is possible that assailants, particularly if they act like or are criminals, could be careless, arrogant, or ignorant about leaving data traces during waging their cyberwar. Another exception is that people within the attacker's camp dare to speak out as whistleblowers despite their personal risks and dangers. Also, knowledgeable human intelligence within or close to assailant's top hierarchy could secretly inform intelligence services about the background of Cyberwar 2.0-related events.

Government's or society's decapitation could start as isolated technical problems. Due to the suppression of certain information, it could take days until these disruptions become apparent to the government or public for what they are. At the same time, rumors, confusion, and intentional misinformation could dominate the information space. Filling an information vacuum does not require Hacker-AI, but it could help to build a coherent narrative.

Even before starting the cyberwar, the assailants could use creative ways to blame others. Within the confusion, they could use uncertainty to arrest people in key positions from bureaucracy, security apparatus, or political class under fake charges while destabilizing the existing order. A carefully planned approach by the attackers is likely more successful than any uncoordinated attempt to resist determined actions. To undermine a country's core institutions, no foreign soldier has to enter the country. Fake evidence about a coup, planted by intimidated collaborators, could be used to destabilize the political system without being immediately detected as a cyber operation.

For technically unsophisticated victims, it is unlikely to recognize Hacker-AI generated malware activities or deep-fakes. Presented explanations for visible cyberevents will be made look more reasonable than a coordinated Hacker-AI-supported cybercoup or cyberwar - until it is too late. Cyberwar 2.0 could have started days before anyone in the affected country could have discovered it. Systematic surveillance of most/relevant people via their electronic devices (smartphones and PCs) could give attackers an uncatchable advantage.

Some security experts think they might have discovered evidence for Cyberwar 2.0, but they could be identified by the attacker in advance and harassed or intimidated by cyberactions. Alternatively, the credibility of these people could be tarnished by having them announce false alarms before. Cyber Reconnaissance and early recruitments of some clerks or officials could trigger these false alarms.

Cyberwar 2.0 assumes that most people (even in relevant positions) can be compelled to collaborate with assailants' demands via direct intimidation by AI bots or real-time deep-fakes in communication or publications. If done inconspicuously, intimidation would not leave any (data) traces, except if some victims have enough courage to report it. Cyberwar is a fast-moving event in which collaborators can't be persuaded; they must be ordered.

Cyberwar 2.0 is likely an undeclared and unannounced war. Suspicious events can be blamed on inherent instabilities in democracies, internal coups, or as side-effects of cybercrime. All Hacker-AI activities are fully deniable because data traces can be removed, or intentional misdirections to installed patsies could be left behind. Even the existence of Hacker-AI will likely be disputed, although the underlying technology is already doable.

With existing cybersecurity tools, cyberdefenders are disadvantaged. Cyberwar activities are covert. At the same time, attackers can follow very complex, detailed plans broken down (and updateable in real-time) to single operational teams with (immediate) feedback on the success/failure of each action step related to their involvement. Even with data traces or records, it is unlikely that we can detect Cyberwar 2.0 actions as a staged event.

The deception of defenders, i.e., misdirection, is part of the plan; otherwise, it would be a lost opportunity. Digital forensics checks data traces, or reverse engineering will find patterns in tool use or other giveaways. Unusual findings in analyzed attack software are attributed to a certain group of hackers. Reusing code snippets, a specific combination of compiler settings/tools, and remnants of the used programming language or character set could all unintentionally reveal information about the attacker, which is then used to attribute authorship.

Hacker-AI could reanalyze developed tools/exploits and create variations that point to different culprits. Besides blaming other countries, it could intentionally create data traces pointing to rogue hacker groups or cybercriminals to deflect responsibilities.

Attacking military supply/logistics and sabotaging weapon release or control systems is considered an act of war - pointing this activity to others is necessary and potentially effective when done systematically with a plan in mind.

Malware from Hacker-AI is potentially detected on servers. There is a chance that honey pots could discover and neutralize these tools. It is also possible that malware is being turned into a double agent to misinform attackers via spoofing. Honey-pots on too many systems are impractical; they lose their value as tools or sources on which their operators could depend for their decision-making. Intelligence could become dangerous if it is part of a trap set out by the assailant to deceive defenders.

The problem for governments facing that kind of adversary is that finding nothing does not mean there is nothing. Instead, it could mean there is something they cannot detect or was not reported.

The new defense line in Cyberwar 2.0 is invisible within the populace. Defenders will intentionally receive plenty of data traces with useless noise or data that should misdirect the defender’s attention or conclusions, but the actual attack is likely undetectable. Without changing this critical deficit/blindness conceptionally, there is little hope that meaningful actions or resistance could alter the outcome of Cyberwar 2.0.

Simulation of Cyberwar Activities

The fog of war is making warfare and every military action a risky endeavor. Many factors remain unconsidered. However, Cyberwar 2.0 actions are likely interactive on small feedback time scales, like 5 or 10 seconds. Then many actions are reversible; there is no regret - Cyberwar 2.0 is destroying nothing. Additionally, simulations could test and study details, outcomes, and consequences. The entire set of capabilities, put into automated war plans/scripts, could show how its strategy/tactics/resources perform under different circumstances. Simulations minimize risks and costs; at the same time, operational tools can be optimized, or their performance can be enhanced with automated follow-up actions. These simulations could run like realistic exercises, including delays from simulated feedback or information/feedback outages.

Cyberwar 2.0 has two levels: strategic and tactical or macro and micro. The strategic level encompasses the big picture, i.e., goals, objectives, and milestones, and is responsible for making decisions and plans that align with those objectives. It is also involved in identifying opportunities and threats. The Tactical level, on the other hand, is more focused on the specific actions and methods that will be used to achieve those goals and objectives, including the evaluation and adjustments of progress. It deals with the details and the execution of the plans. Therefore, Cyberwar 2.0 has two corresponding war plans/scripts.

The tactical war script will deal with the automated responses for situations expected within the interactions with collaborators. The tools using this script could still be designed to ensure that significant cyberwar actions are controlled by humans proactively - if required.

The Cyberwar 2.0 software platform could track success and adjust to failures (automatically). Instead of destroying enemy capabilities permanently, temporary/remotely-controllable acts of sabotage or service denial can be prepared, tested ahead of their activation, and deactivated or modified if required.

Most importantly, realistic simulations are based on actual data from detailed reconnaissance and surveillance; near-real-time responses and feedback can be simulated with probabilistic models - indicating that sometimes plans fail. Within these simulation models, geography, population details, asset distributions or deployments, possible adversarial threats, etc., could be considered. The simulation could be played like a game in which the assailant's war script is challenged with unexpected events or coordinated resistance.

Additionally, the impact of sanctions on the economy is a likely a topic of simulation that detects vulnerabilities proactively. Some countries could use simulations to determine how preventative espionage measures ahead of an attack could reduce sanctions' impact on their economy.

The lesson from the Ukraine war is that countries intending to start a war of aggression should be much better prepared for sanctions. The same applies to the world community; countries will likely know and understand much better what sanctions they could issue and how their vulnerabilities or dependencies could be reduced by preparation.

Hacker-AI-generated spyware can be used against key-supply companies to gain the know-how to prevent sanction-triggered business continuity disruption. This proactive approach to sanctions via targeted espionage is not new.

However, Hacker-AI could be used to apply it more precisely. This is likely another net-positive contribution of using Cyberwar 2.0 and Hacker-AI for the assailant and the targeted country's economy to reduce the cost of the war even in its aftermath.