ISSN: 2663-5798 || Arab Journal for Scientific Publishing

المجلد التاسع – العدد الخامس والتسعون
تاريخ الإصدار: 2 – أيلول – 2026م

www.ajsp.net

"HL and AI Cyberattacks"

Researcher:

Diana Al-Hajj Sulaiman

Received: 09/08/2026 | Revised: 10/08/2026 | Accepted: 18/08/2026 | Published: 02/09/2026

Abstract:

The article explores the effectiveness of IHL in the context of effectively regulating cyberattacks supported by AI. The research examines and analyzes the principles of IHL—distinction, proportionality, precaution, and accountability—in relation to a standalone operation of CI. This study is a doctrinal-legal collection of the existing treaty frameworks and customary international law and of foundational documents like the Tallinn Manual 2.0. The analysis reveals three general patterns: IHL has been conceptually extended to cyberspace, compliance gaps exist in the context of autonomy, and a gap in accountability exists. The results indicate that while the current IHLs extend to cover cyber operations, the involvement of autonomous AI systems calls into question the adherence of IHLs. The article concludes that the market is not inherently bad, current legal tools are not sufficient, and that a combination of treaty adaptation and legally binding treaty instruments specifically launched for autonomous cyber weapons would be desirable. Real-life case studies of Stuxnet, the Gaza AI targeting controversy, and the attack on Saudi infrastructure emphasize the regulatory need for reform.

How to Cite This Article

Al-Hajj Sulaiman, D. (2026). IHL and AI cyberattacks. *Arab Journal for Scientific Publishing (AJSP)*, 9(95), (428-437).

AJSP | Vol. 9 | Issue 95 | DOI: https://doi.org/10.36571/ajsp.95 **AJSP ORCID: https://orcid.org/0009-0005-8048-2082**

Introduction:

Digital warfare is described as an increasing characteristic of modern warfare. The provision of advanced 'cyber weapons' by states to disable, disrupt, and ultimately destroy enemy capabilities, without resorting to kinetic weapons or war. The involvement of artificial intelligence (AI) in such cyber warfare operations is a game-changer and something rather unsettling about the nature of armed conflict. The attacks are automated and can have physical repercussions on critical civilian infrastructures delivered at machine speed without direct human consent at the time of the attack.

Basic principles of human decency, or International Humanitarian Law (IHL), have always set rules as to the means and methods of armed conflict, imposing on those who have waged wars obligations to respect them. IHL is fashioned at a time when men and women fought with and in a human army and with actual weapons and battle lines. Now comes the burning but unanswered question: can IHL adequately control cyberattacks using AI tools? This article contends that, in essence, the rules of IHL are poorly fitted to autonomous AI systems, though they may be technically applicable to cyber operations. The distinction, proportionality, and precautions are seriously challenged in the cyber world. Additionally, the legal framework is not in place to adequately address this accountability gap.

The remainder of this article will be presented in two parts. Part I reflects on the legal significance of applying IHL in a cyber context, such as whether AI-cyber operations are means and methods of warfare and whether the very principles of IHL apply. Part II offers an analysis of legal challenges arising from autonomous AI, such as the accountability gap and the need for novel regulatory structures. Three Middle East case studies (Stuxnet, Gaza targeting through AI systems, and Saudi Arabia's Shamoon attacks) are used as a thread through the book to illustrate the analysis.

ISSN: 2663-5798 || Arab Journal for Scientific Publishing

المجلد التاسع – العدد الخامس والتسعون
تاريخ الإصدار: 2 – أيلول – 2026م
www.ajsp.net

Part I: The Applicability of International Humanitarian Law to AI-Driven Cyberattacks

Section 1: The Legal Recognition of Cyber Operations Under IHL

A. Cyberattacks as Weapons and Warfare

The key question is whether cyberattacks are means or methods of fighting a war under IHL. Conventional sources of IHL, such as the Hague Conventions and the Geneva Conventions, were drafted primarily in relation to the use of physical weapons and kinetic operations. The ICRC has always believed that IHL is applicable to all kinds of war, including the technologically led ones (ICRC, 2019). Instead, IHL has been broadly interpreted to include digital operations that have effects similar to conventional attacks.

Article 35 Additional Protocol The 1977 Hague Law on the lawful conduct of war bans means and methods of warfare of a nature to cause superfluous injury or unnecessary suffering. It can totally apply to cyber weaponry if it has indiscriminate, irreversible, or disproportionately bad effects. If it is a weapon specifically created to disrupt industrial control systems (DC), for instance, intending to inflict unnecessary suffering, then it is a violation of the prohibition. Article 36, banning new weapons incompatible with the laws of war, also applies: States are obligated to ensure that new cyber weapons are legally examined before use.

The protection of people in armed conflict under the Geneva Conventions also apply. The Geneva Convention IV & Additional Protocol I provides for the protection of civilians and civilian objects. Under Article 52(2) of Additional Protocol I, attacks are to be directed only against military objectives, which, by their nature, location, purpose, or use, serve military action or its potential. This raises more concerns about whether cyberattacks against infrastructure with both military and civilian consumers satisfy this definition (Mazal, 2024). The Tallinn Manual 2.0 has reiterated the definition of a cyber-attack as an 'attack' under IHL if it is 'reasonably likely to cause' injury, death, damage, or destruction.

AI-powered cyberattacks significantly exacerbate these worries. AI-powered cyber technologies can be set up to adapt their tactics on their own, searching for networks in ways never coded into them. The United States and Israel reportedly created Stuxnet, and it has been used to launch attacks against the Iranian nuclear centrifuges at the Natanz facility, making it the most sophisticated and advanced offensive cyber weapon ever deployed (Umar, 2025). Stuxnet, although not necessarily AI-based, was a self-propagating cyber malware that explored the legal problems raised by full-fledged autonomous cyber malware. So future killer robots with AI could be able to find and kill industrial systems without requiring much human oversight.

B. Conditions for the Application of IHL

IHL is applicable to armed conflict only, according to Common Articles 2 and 3 of the Geneva Conventions. IHL is applicable to two situations of armed conflict: international armed conflict (IAC) between two or more states and non-international armed conflict (NIAC) occurring in the territory of one state. Legally and practically speaking, there remains a debate whether a cyberattack can or should create an armed conflict or whether it has to occur during an armed conflict.

For any IAC to be legitimate in the context of cyber operations, its scope needs to cross above a threshold like conventional armed attacks. According to the Nicaragua Case (1986) of the ICJ, an armed attack needs to be "substantial" and must be "on a large scale," criteria that most cyber incidents do not satisfy. That could all go out of the window with cyber operations designed to cause physical damage, including systems targeted against power grids or water systems. In the Tallinn Manual 2.0, Schmitt (2017) contends that it is not the means used that are of relevance, but the effects obtained. If a cyber-attack causes physical damage comparable to conventional military attacks, it constitutes an armed attack and falls within IHL.

The Gaza context has seen cyber activity being operated alongside kinetic military activity, placing it within the context of a current armed conflict, in which IHL clearly applies. Israel is one of the states that has been using AI on the battlefield during its military operations in Gaza, where it uses AI-powered targeting systems. They took place during an armed conflict. They were covered by the IAC provisions of Additional Protocol I, as well as by the duties imposed on military commanders under IHL to take all feasible precautions in an attack.

IAC and NIAC are different in terms of the rules that apply. The rules that governed hostilities in particular detail—Additional Protocol I—apply only to IACs and to wars of national liberation. Common Article 3 and Additional Protocol II (only applicable in non-international armed conflict) do not contain specific rules regarding the conduct of hostilities. The significance of this gap lies in the fact that AI-driven cyber operations are employed by non-state actors or in mixed conflicts, as is becoming more prevalent in the Middle East.

International humanitarian law (IHL) conventions need to be updated to accommodate the new reality of AI-driven cyber warfare. Collectively binding treaties must be written by the states to address the use of independent cyber weapons in wartime. The veil of indetermination in the law will continue to cast shadows of accountability that jeopardize the protection afforded to civilian populations at large. AI systems with independent decision-making abilities pose a challenge to the fundamental IHL principle of meaningful human control over attacks. When choosing military targets, great care must always be taken to distinguish between civilian objects and real military targets—this is the principle of discrimination. Autonomous systems are launching attacks, and the human commander does not even evaluate them. Article 57's precautionary measures should be further considered to encompass algorithmic decision-making procedures in cyber operations. An AI cyber weapon must be legally reviewed and certified nationally compliant with Article 36 of the Constitution prior to deployment. The development of technical standards for the design of autonomous cyber tools should include the involvement of international bodies, such as the ICRC. This failure would allow ignorance and negligence to give way to unacceptable AI warfare with far-reaching and lasting global humanitarian consequences.

Section 2: Applying Core IHL Principles to AI Cyber Operations:

A. Distinction and Proportionality

Distinction is a principle. Article 48 of Additional Protocol I to the Geneva Conventions: In the case of an armed conflict, the parties hereto shall, always, distinguish between civilians and combatants, between civilian objects and military objectives. In attack mode, only combatants and military targets can be attacked. The presence of AI in cyber-attacks introduces interdependent complexities to this fundamental rule, underlining structural difficulties in applying IHL to the modern cyber war environment.

Furthermore, today's civilian and military networks are tightly interconnected. Military command and control is built on the same digital infrastructure used for electricity, telecom and water treatment, financial networks, and hospital systems. The nature of the cyber-attack may require traversing and impacting the civilian network from the military communications node. autonomous AI systems may lack the 'situational awareness' to know which portion of this shared infrastructure is a combat target at any given time (Roscini, 2014). Unlike a human commander making battlefield decisions, the AI's training data and programmed criteria may not involve the nuance of civilian and military infrastructure.

Secondly, a potential misuse by AI of a target has serious implications for the distinction principle. The machine learning algorithms used in targeting systems are based on experiences of the past, which might include bias or could fail to extrapolate to the new situation. For the current episode in Gaza, Israeli sources report the use of an AI-powered targeting system called 'Lavender,' which algorithmically combined surveillance data to generate lists of potential targets (Haaretz, 2024). Those targets were supposed to be produced by the system in tens of thousands, with varying degrees of confidence, and yet there were still significant question marks about the judgment exercised by the human operators to allow them to be targeted. Likewise, AI systems instructed to detect and target enemy networks might mistakenly label civilian networks as targets.

Thirdly, the ripple effect of cyberattacks on critical infrastructure is an immediate issue in terms of distinction and proportionality. The proportionality principle allows for an attack that would result in civilian injury only if an expected military benefit has been considered. AI-driven attackers are able to exploit interconnected networks, where the number of casualties may not be definitively known at the time of an attack. Stuxnet, for instance, a computer worm designed to control specific computer systems at Natanz, also managed to affect unintended civilian systems in several other countries (Umar, 2025). Stuxnet, with an "AI-powered" capability, has the potential to spread much faster and farther, making calculations of proportionality seem quaint.

The Saudi Arabian Shamoon attacks are a case in point of how unsecured civilian critical infrastructure can be. As part of the 2012 attacks on Saudi Aramco, it is estimated that the Shamoon malware damaged some 35,000 PCs and was backed by Iran (Kausch, 2022). Aramco is state-owned but is a part of the civilian economic infrastructure. A computer-controlled version of Shamoon could be mass-distributed across networks throughout the energy sector and cause devastation to oil supply, economic collapse, and shutdown of infrastructure on which civilians also depend. The proportionality of any such attack would depend on almost exact knowledge of the lines of communication and the extent of civilian reliance, which a military commander could not reasonably be expected to possess.

B. Precaution and Military Necessity

The provision of Article 57 of the Additional Protocol I imposes a duty of caution upon belligerent forces to take "all feasible precautions" to avoid or to reduce civilian casualties. These obligations encompass ensuring that the targets are

military objects, selecting means and methods that will minimize incidental civilian harm, and ceasing or suspending attacks when incidents are likely to cause harm to civilians that is disproportionate. The application of the precautionary principle is quite problematic in the case of autonomous AI cyber systems in terms of predictability, autonomy, and human control.

This predictability problem is underlying. AI systems, in particular deep learning architectures, are known to be unpredictable in new environments. An apparent 'black box' problem is when even the creators of the decision-makers do not understand the rationale behind a specific selection. Unpredictable weapon effects may inhibit potential attackers from accounting for incidental damage and precautionary responsibilities. An autonomous cyber weapon deployed against an adversary's network might adapt in unexpected ways, encountering systems, security controls, and network configurations never previously encountered.

These concerns are compounded by the fact that autonomous malicious software is designed that way. Traditional malicious software with known behavior can be designed to be self-contained, evolve to its environment, self-replicate, and change attacking vectors. Once operational, such weapons might be able to selectively track and shoot targets independently, without human authorization, on a target-by-target basis. The ICRC has stated that these types of weapons are incompatible with IHL and can only be used if a human being would actually declare to employ them, taking into consideration each military and humanitarian circumstance of their application (ICRC, 2021). The distinction principle and proportionality rule cannot possibly be fulfilled on a case-by-case basis without providing any meaningful human control.

Control by humans is therefore a matter of IHL and not good practice. In international debates, the principles of 'meaningful human control' have been used as a measure for an autonomous weapon system. Humans must be empowered with real capacities to make choices about targeting human life and death, and there must be the capacity to understand, predict, and interfere with the operation of the system, according to this criterion. AI-controlled cyber weapons posing a threat and deployed in nanoseconds without human decision-making, hard-coding of algorithms, and proliferation of weapons until the human player adheres to the rules do not fit this criterion (Article 36, 2015). The loss of control over the use of force is not merely a technical flaw but also a structural failure in complying with the precautionary requirement in IHL.

Part II: The Legal Challenges Created by Autonomous AI Cyberattacks

Section 1: The Accountability Gap

A. Difficulties in Attribution

One of the biggest challenges with regulating AI-assisted cyberattacks is identifying the origin of the attack. International law imposes legal responsibility, calling for attribution. In the absence of attribution, it is impossible to hold people or states accountable for actions that cause harm. The rise of the AI-powered cyberattack gives added complexity to attribution challenges and an added threat of a zone of impunity.

The Internet itself and its use for cyber activities prevent claims of attribution since attackers may use commercial cyber infrastructure and sophisticated techniques of obfuscation and be able to operate across multiple jurisdictions. State-sponsored cyber activity is frequently done through proxies, non-state actors, criminal organizations, or front companies that provide plausible deniability. Stuxnet, for example, was a years-long attribution to the U.S. and Israel and never officially confirmed, either by the U.S. or Israel (Umar, 2025). AI-driven cyber operations make the attribution process tougher due to their ability to adapt, evade, and overcome traditional detection techniques.

In addition to this, AI decision-making introduces a layer of complexity to the process of attribution, which might be a black box operation. The harmful outcomes of autonomous AI make it hard to attribute them to either programming, emergence, or manipulation. Misdirection could be a way to divert an AI cyber weapon to civilian infrastructure that they had planned to leave untouched in the case of an adversarial attack (Luthar, 2025). In that case, determining which party was at fault is extremely complicated.

The relevance of state responsibility in international law is best summarized by the Articles on Responsibility of States for Internationally Wrongful Acts (ARSIWA), adopted in 2001 by the International Law Commission (ILC). Article 4 of ARSIWA recognizes the presence of acts of state organs as 'acts' of the state. Acts of non-state actors will be attributed to a state under Article 8 if they were carried out under the instructions or on the direction or under the control of the state. The confiscation of AI abuses and arbitrary decisions can make it harder to determine whether an attack is a result of a state action or an AI decision. This ambiguity results in a possible accountability loophole, damaging the victims of such attacks (ILC, 2001).

The Iran–Israel cyber conflict is a telling example of these attribution issues. Israel has been accused of watering down the water systems of Iranian cities and causing water infrastructure disruptions in a number of cities (Times of Israel, 2020). Iran, in turn, has been accused of attacks against Israeli water treatment plants, targeting increases in levels of chlorine

in water to unsafe levels. Both states denied responsibility for their activities. An AI that operates independently and with full autonomy makes it hard to identify and overlook the fingerprints of human operators.

B. Individual Criminal Responsibility

IHL, and also international criminal law, make people who provide criminal orders or who are neglectful to prevent war crimes criminally liable. War crimes committed during armed conflict are part of the crimes under the Rome Statute of the International Criminal Court (ICC). Orders, requests, and facilitating the commission of war crimes are also criminal acts under Article 25, par. 2 of the Rome Statute of the ICC. Military commanders are responsible for offenses by troops under their effective command, of which they had, or should have had, knowledge and for which they took, or could have taken, no steps to prevent or punish such offenses pursuant to Article 28.

AI-powered cyberattacks present deep conceptual and practical issues when it comes to command responsibility under international law. In the case of command responsibility, several things need to be achieved, including that the commander must have exercised 'effective control' over the crime committed by the subordinates. An original concept in the electronics industry was that of "good control," in which people who are lower than you in rank are hired as part of a system of commands. This type of delegation of authority to an AI system executing a 'fire' order in every state without the human reaction time would have to be addressed in terms of the AI making autonomous decisions.

The Gaza war is an example of such targeting by AI. According to some reports, a system known as 'Gospel' activated and circulated targets to be struck, including by Israeli military units, in a matter of minutes after the AI had created the recommendation (+972 Magazine, 2024). The vast scale and speed at which AI can make targeting recommendations could diminish the role of human oversight in relevant command decisions. With targeting suggestions being generated quickly, human operators may not have enough time to make meaningful decisions. Under current IHL, it is unclear as to what the answer would be because it was not envisioned for such an autonomous human-machine teaming.

The responsibility gap refers to the complexity in determining who should be held liable when AI systems function autonomously and result in unanticipated harm or consequences. The damage can occur without permission or human guidance and supervision (Osman et al., 2025). Therefore, the current legislation would make it difficult to decide who would be considered a criminal according to the legal system. A specific harmful action is not the programmer's intention when implementing the AI. The status of the launcher may not be aware of the target. The political decision-maker who issued an order to use AI is liable for the overall use of AI. This division of responsibility among the different human actors involved, and none of them, creates a legal void that militates against the deterrence function of IHL.

Section 2: Towards New Regulation of Military AI

A. Limits of Existing Legal Frameworks

Existing IHL frameworks are now known to be poorly suited to regulating cyberattacks by AI systems. The Geneva Conventions and Additional Protocols were drafted in 1949 and 1977, well before digital networks could be dreamed of and automated weapons existed. The Martens Clause also offers some protection as a residual clause (ICRC, 2021). However, it cannot substitute the individual rules specific to the unique challenges AI faces in cyber warfare.

Further, the LCOE at NATO gathered an international group of law experts to draft the Timbuktu Manual 2.0, the most systematic effort to date to apply existing IHL to cyber operations. The manual indicates that the underpinning principles of IHL apply to cyberattacks and sets out various specific rules to be applied. Moreover, important issues exist on which the law is yet to be clarified, even according to its own manual (Schmitt, 2017). However, the Tallinn Manual is not legally binding but rather a reflection of the views of individual experts and a non-confirmation of new treaty law.

Since 2014, the Convention on Certain Conventional Weapons (CCW) has been in a talking session about lethal autonomous weapon systems (LAWs). A consensus has emerged out of these discussions that the application of autonomous weapons under IHL must contain aspects of human control. The CCW process has so far failed to secure a legally binding treaty on autonomous weapons. Moreover, a key focus of the CCW process is kinetic weapons, with the Cyber CBRN developments underdeveloped (UN GGE, 2023).

This is like autonomous drones and autonomous cyber weapons. Both make the same kind of distinction: proportionality and accountability inquiries. Cyber weapons are invisible and uncontrollable and barrage their targets many times. Beyond IHL, no specific legal regime for digital 'land mines,' in this case self-propagating malware, exists. There is no international agreement on an obligation to "disable or neutralize" ACEs at the end of hostilities, as is the case with mines under Protocol II of the CCW.

B. Options for Reform: Adaptation or New Treaty

Legal interpretation and application of IHL as it stands will fill in the regulatory gap, and so will a dedicated treaty for cyberspace military AI via treaty negotiations. Each of them has advantages and disadvantages that need to be taken seriously in view of the threat of the regulatory problem and the pressures of international treaty negotiations.

First, the adaptation of existing IHL assumes that the basic rules are sufficiently drawn up to contain new technologies in an "adaptive" interpretation of the rules themselves. An additional Protocol I article requiring implementation of 'human control in the meantime on targeting' (Article 57). Although it is not explicitly a decision-making process in controlling language, it is an implicit one still controlled by humans. States might adopt authoritative interpretations of current treaty language that include explicit provisions for interoperability, such as tests for algorithmic transparency, pre-deployment testing, and the involvement of human actors in real time. The ICRC suggests requirements that must be feasible for any autonomous weapon system to be limited in (Miniak, 2021).

The second option, a new binding treaty, is grander yet may be necessary given technological advances. A key challenge is that AI operations in military contexts can be governed by a separate treaty dealing with the particular problems of autonomous systems in cyber warfare. Such a treaty could also include a floor of human oversight, restrict the use of lay-up technology as an inherently inhuman weapons system, and define standards for responsibility and liability for human harm committed by AI. The Chemical Weapons Convention and Biological Weapons Convention are examples of treaties that place a categorical ban on purely indiscriminate arms. Today, there is a theoretical possibility of prohibiting cyber weapons that have the capacity to be completely autonomous and would not be distinguishable or proportionate. However, there is political opposition because these independently operating weapons benefit power-strengthening states militarily.

Both options can be encouraged through multilateral forums via the UN processes. Responsible state conduct and the upholding of IHL have been advocated by UN expert bodies and working groups in cyberspace. The lack of these rules has been impeding a legally binding nature due to geopolitical issues between the actors. The Gaza and Iran-Israel wars reveal the need for maneuvering space for independent cyber actions all over the world. Politically challenging and legally critical in regard to maintaining humanitarian standards for the protection of civilians.

Current law contains many gaps and weaknesses regarding fundamental aspects that are exposed in the scenario of autonomous AI warfare. The overall poor track record of state actors in turning their diplomatic words and actions into legally binding terms that cover the military application of AI. The world is now in dire need of more enabling mechanisms than simple voluntary norms and expert manuals without binding force. Substantial progress towards autonomous offensive cyber weapons will be a challenge to achieve without the political reticence of the big military powers. The dangerous liberty of powerful states, beyond the law, is the common use of AI systems against civilians. The development of AI technology grows much faster than the development of international law, leaving room for regulatory loopholes at all times. This created a badly fragmented legal system that was conducive to aggressive nation-states and not to protecting vulnerable civilian populations. It is high time for the states to realize they must develop new, binding legal frameworks for autonomous AI cyborg weapons. Multilateral negotiations must be accelerated if autonomous AI cyber systems are to be the norm, rather than the exception, in the violation of fundamental humanitarian principles. International law should not be pushed aside as more and more autonomy is granted to machines that are required to govern parts of human life.

Case Studies: AI Cyber Warfare in the Middle East

Case Study 1: The Iran–Israel Cyber Conflict and Stuxnet

The poster child of a cyberattack causing real-world damage is the 2010 worm Stuxnet. Stuxnet hit Siemens programmable logic controllers and disrupted about 20% of Iran's nuclear centrifuges at Natanz. It caused centrifuges to run at erratic speeds and falsely indicate they were operating normally on all monitoring systems. The worm was very complex, contained a number of zero-day exploits, and targeted only certain industrial settings. Despite Stuxnet's accuracy, it went on to infect civilian systems worldwide, including Indonesia, India, and elsewhere.

Stuxnet asks a basic series of questions about IHL regarding the principles of distinction between civilians and combatants and proportionality. The Natanz facility has a dual purpose and is used in Iran's nuclear program, which has civilian energy and possible military aspects. Additional Protocol I, Article 52(2), requires the following: the destruction of the facility must provide a "definite military advantage "this is in dispute. The idea of proportionality becomes more complicated if starting a war over Iran's nuclear option is militarily beneficial to Israel, if not now, then certainly, but not necessarily legitimate. The worm was deployed and caused unanticipated collateral damage as it propagated across systems around the world.

After Stuxnet, cyber warfare between Iran and Israel has become a fight on water, ports, and financial networks, as well as all critical infrastructure in Iran. In 2020, Israel is reported to have attacked the Iranian town of Bandar Abbas; the Shahid Rajaee port, which has disrupted the port's activities and been costly to the civilian commercial trade sector (Washington Post, 2020). According to the report, the 'reciprocal' attack by Iran on Israeli water infrastructure targeted contaminating water supplies of the civilian population. Any targeting of civilian objects by AI is obviously a violation of IHL. Additional Protocol I, Article 54, also includes provisions on precautions regarding the protection of objects indispensable to the survival of the civilian population.

Case Study 2: Gaza and AI-Enabled Targeting Systems

Gaza AI targeting is a trial run for how IHL can control military AI. After the Gaza war of 2023–2024, investigative journalism outlets report that the Israeli military used AI systems known as 'Lavender' and 'Gospel' to recommend targets. According to the Lavender system, tens of thousands of individuals were identified through digital surveillance, analysis of communications metadata, and social media communications as potential targets (Haaretz, 2024). Human operators were said to have virtually no time to review AI recommendations before authorizing a strike.

This raises relevant questions about IHL. The doctrine of distinction requires that, before an attack can be permitted, the target be individually specified as a combatant or military objective. AI target lists could only have been superficially checked by human operators if there was no independent evaluation. Under Article 57(2)(a)(i) of Additional Protocol I, a positive duty of diligence to determine whether a specific target is a military objective may not be fulfilled by rubber-stamp adoption of AI recommendations. Targeting systems based on AI can include algorithmic biases and remain invisible offenders of distinction.

Other aspects of gender, civilians, and structural discrimination in armed conflict and AI targeting are also highlighted in the case of Gaza. In Gaza, for example, because women and children comprise a large civilian population, the use of AI in military operations targets these vulnerable groups. The difficulty of autonomous targeting systems to recognize and protect their status is an issue of concern as mass casualties occur. The ICRC has emphasized that AI targeting systems must recognize and respect the protected status of civilians, medical staff, and humanitarian workers (ICRC, 2021).

Case Study 3: Saudi Arabia and the Shamoon Attacks

The Shamoon cyberattacks against Saudi Aramco in 2012, and then Shamoon 2 in 2016-2017, are the most destructive cyberattacks to date against critical civilian infrastructure. Shamoon caused the master boot record to be destroyed on many—estimated to be tens of thousands—of computers, causing them to become unusable. In 2017, multiple Saudi government ministries and energy companies were targeted, and critical services were unavailable for an extended time (Ouassini & Boynton, 2021). The attacks were thought to be launched as part of Iranian state-sponsored operators' retaliation effort for Stuxnet and other attacks against Iran.

The Shamoon attack is just one such instance of the vulnerability of civilian critical infrastructure and the laws that should be in effect. Under Article 52(1) of Additional Protocol I, civilian objects are not subject to attack or reprisals. Saudi Aramco is state-owned but operated as a civilian economic infrastructure, and it is vital to the supply of energy to civilians. The Shamoon variant is self-propagating, capable of being disseminated via energy networks, and causing civilian damage that they could not otherwise have caused. Meanwhile, it significantly increases the difficulty of tracing the origins and responsibility for these AI enhancements in impacted states.

The Shamoon cases also bring up the issue of reprisals. Article 51(6) of Additional Protocol I state that the "civilian objects" must not be targeted by violators of IHL even if Iran believed that they could be safeguarded as "belligerent reprisals" for violations of IHL committed by Stuxnet. A reprisal attack by AI against civilian infrastructure is doubly illegal, as it is an attack on civilian infrastructure as well as being a disproportionate reprisal. One of the more alarming emerging scenarios is the use of AI systems for autonomous means to attack the civilian infrastructure of the enemy as reprisal operations. Such military automated acts of retribution cannot be prevented and/or effectively controlled by the currently existing IHL regimes.

Conclusion

This article has maintained that International Humanitarian Law (IHL) applies to cyberattacks using artificial intelligence. However, it is poorly suited to regulate the use of artificial intelligence in warfare effectively. An autonomous artificial intelligence system deployed in cyberspace raises the question of whether the general principles of distinction and proportionality and precaution apply to this in practice. IHL's human command focus is fundamentally at odds with the rapid, opaque, volatile reality of autonomous AI in operation.

The most important is the lack of accountability. Issues of deterrence and justice are affected by the absence of accountability for the programmers, commanding officers, and political authorities, as well as the 'black box' of the artificial intelligence decision-making process. Victims of AI cyberattacks suffer from incomplete and uncertain remedies based on

the notions of state responsibility (under ARSIWA) and command responsibility (under the Rome Statute). The Middle East case studies illustrate that these are no hypothetical concerns: Stuxnet is already a reality and is associated with attacks on targets in Gaza. At the same time, Shamoon is being used in attacks against targets.

Two ahead tracks are found. If IHL is interpreted to include the requirement for meaningful human control to be a prerequisite for, or a condition of, autonomous systems' compliance with IHL, evidence of meaningful human control will be a core element of autonomous systems' compliance with IHL. Because meaningful human control is a key issue for AI cyber weapon systems, it is essential that clear standards be developed by states, the ICRC, and international law experts on what is meant by "meaningful human control" in these systems. Second, a lawfully binding international agreement is required that is dedicated in nature to the goal of controlling autonomous cyber weapons and will not be subject to interpretation. This treaty should prohibit autonomous cyber weapons incapable of respecting the law of war, define attribution of any damages resulting from the use of AI, and set out accountability requirements for states and individuals. The urgency of this need is higher than ever before as the momentum and uptake of AI technology are accelerating across the globe, and the runway for successful international regulation is fast closing.

References:

+972 Magazine. (2024). Lavender: The AI machine directing Israel's bombing spree in Gaza. https://www.972mag.com/lavender-ai-israeli-army-gaza/

Alfatlawi, A. A., Tamimi, K., & Al-Tamimi, M. J. A. (2024). Feasible precautions: A legal study in the stable and variable concept under international humanitarian law. Journal of the College of Law for Legal and Political Sciences, 10(37), 598–629. https://papers.ssrn.com/sol3/Delivery.cfm?abstractid=4957708

Article 36. (2015). Killer robots: UK government policy on fully autonomous weapons [Briefing paper].

Haaretz. (2024). Lavender: The AI system the IDF used to target 37,000 suspected Hamas operatives in Gaza.

International Committee of the Red Cross. (2019). International humanitarian law and cyber operations during armed conflicts: ICRC position paper.

International Committee of the Red Cross. (2021). ICRC position on autonomous weapon systems.

International Court of Justice. (1986). Military and paramilitary activities in and against Nicaragua (Nicaragua v. United States of America). ICJ Reports.

International Criminal Court. (1998). Rome Statute of the International Criminal Court. United Nations Treaty Series, 2187, 3.

International Law Commission. (2001). Articles on the responsibility of states for internationally wrongful acts (ARSIWA). United Nations.

Kausch, K. (2022). Cheap havoc: How cyber-geopolitics will destabilize the Middle East. German Marshall Fund of the United States. https://www.gmfus.org/sites/default/files/Cheap%2520Havoc%2520-%2520How%2520Cyber-Geopolitics%2520Will%2520Destabilize%2520the%2520Middle%2520East%2520edited.pdf

Martins Luthar, A. E. (2025). Adversarial AI: When hackers use artificial intelligence against security systems. https://www.researchgate.net/profile/Rapheal-Alamu/publication/396843225_Adversarial_AI_When_Hackers_Use_Artificial_Intelligence_Against_Security_Systems/links/68fabf68e7f5f867e6e1f1e2/Adversarial-AI-When-Hackers-Use-Artificial-Intelligence-Against-Security-Systems.pdf

Mazal, J. (2024). The dual use of civilian and military technologies on the battlefield of the future. https://real.mtak.hu/210721/1/255-308.pdf

Osman, Y. H. M., John, D. J., Ahmed, D. A., Mohamed, R. M. A., & Ali, R. M. S. A. (2025). Criminal responsibility in the age of AI: Rethinking legal accountability. African Journal of Law and Justice System, 4(3), 179. https://journals.co.za/doi/pdf/10.31920/2753-3123/2025/v4n3a9

Ouassini, A., & Boynton, K. W. (2021). The "Silicon Valley of the Middle East": Cybersecurity, Saudi Arabia, and the path to Vision 2030. In Routledge companion to global cyber-security strategy (pp. 427–434). Routledge. https://www.researchgate.net/profile/Scott-Romaniuk/publication/347767051_Routledge_Companion_to_Global_Cyber-Security_Strategy/links/620a2649cf7c2349ca12e87c/Routledge-Companion-to-Global-Cyber-Security-Strategy.pdf#page=451

Roscini, M. (2024). International law and the principle of non-intervention: History, theory, and interactions with other principles. Oxford University Press.

Schmitt, M. N. (Ed.). (2017). Tallinn manual 2.0 on international law applicable to cyber operations. Cambridge University Press.

The Times of Israel. (2020). Report: Israel carried out a cyberattack on an Iranian port in retaliation for an infrastructure hack.

Umar, M. (2025). Iran and Israel's cyber warfare and interference in the US. Social Science Review Archives, 3(2), 2040–2048. https://policyjournalofms.com/index.php/6/article/download/824/838

United Nations Group of Governmental Experts. (2023). Report of the Group of Governmental Experts on emerging technologies in lethal autonomous weapons systems (UN Document A/78/XXXX).

The Washington Post. (2020). Officials: Israel linked to cyberattack on Iranian port.