Regulating Digital Weapons: Autonomous Cyberattacks, AI Warfare, and International Humanitarian Law

1. Liam Patterson: Department of Criminal Justice, University of Sydney, Sydney, Australia
2. Mariana Scouza*: Department of Political Science, University of São Paulo, São Paulo, Brazil

***Correspondence: e-mail: mariana.scouza@usp.br

Abstract

The rapid integration of artificial intelligence into cyber operations has transformed the nature of contemporary conflict, producing digital weapons capable of autonomous decision-making, adaptive targeting, and machine-speed escalation. These developments challenge long-standing assumptions within International Humanitarian Law (IHL), exposing doctrinal gaps that traditional legal frameworks are not yet prepared to address. This narrative review examines the technical foundations, operational dynamics, and legal implications of autonomous cyberattacks and AI-enabled warfare, synthesizing insights from technology studies, security analysis, and humanitarian law. The discussion begins by contextualizing the emergence of digital weapons, defining autonomous cyber systems and AI warfare, and outlining their growing relevance in military doctrine and strategic competition. It then analyzes the challenges these systems pose for IHL, particularly regarding distinction, proportionality, necessity, and precaution, as well as attribution, foreseeability, dual-use infrastructure, and the ambiguity surrounding data as an object of attack. The review further evaluates existing regulatory approaches, including the Tallinn Manual, UN cyber governance initiatives, regional frameworks, and national strategies, highlighting their limitations in addressing autonomous escalation, opaque algorithms, and self-learning cyber capabilities. Building on this analysis, the final section proposes foundational elements for a coherent legal and ethical framework that integrates meaningful human control, establishes clear accountability mechanisms, promotes shared definitions of autonomous cyber weapons, strengthens due-diligence obligations, and embeds ethical principles into AI system design. The article concludes that governing digital weapons requires innovative regulatory models that combine legal, technical, and ethical expertise. Only through coordinated international efforts can states ensure that AI-enabled cyber operations evolve in ways that preserve humanitarian protections, enhance accountability, and promote stability in an increasingly digital battlespace.

Keywords: Autonomous cyberattacks; AI warfare; digital weapons; International Humanitarian Law; cyber conflict; algorithmic accountability; meaningful human control; cybersecurity governance; autonomous weapons regulation; cyber norms.

Received: date: 12 February 2023
Revised: date: 12 March 2023
Accepted: date: 26 March 2023
Published: date: 01 April 2023

Citation: Arslan, S., & Scouza, M. (2023). Regulating Digital Weapons: Autonomous Cyberattacks, AI Warfare, and International Humanitarian Law. *Legal Studies in Digital Age*, 2(2), 49-60.

1. Introduction

The landscape of contemporary warfare has undergone a profound transformation as states and non-state actors increasingly rely on digital technologies to project power, challenge adversaries, and defend strategic interests. The shift from conventional, kinetic forms of conflict toward technologically enhanced and automated operations is driven by rapid developments in artificial intelligence, machine learning, and autonomous systems. These technologies have given rise to what many scholars describe as "digital weapons," a broad category that includes AI-enabled malware, autonomous cyber intrusion tools, and self-learning systems capable of identifying vulnerabilities, selecting targets, and executing operations with minimal human oversight. The emergence of digital weapons has become a defining feature of twenty-first-century conflict, reshaping military doctrines and altering the dynamics of international competition as states develop increasingly sophisticated offensive and defensive cyber capabilities in response to evolving threats, shifting geopolitical tensions, and the strategic advantages conferred by superior technological power.

Digital weapons can be understood as software, algorithms, or cyber-physical hybrids designed to infiltrate, manipulate, degrade, or destroy digital systems or critical infrastructure. Researchers examining cyber conflict have noted that contemporary cyber operations often involve highly adaptive malware, automated network reconnaissance tools, and autonomous intrusion mechanisms that can respond to environmental conditions in real time, thereby complicating defensive efforts and increasing the risk of uncontrolled escalation (Tang et al., 2023). Autonomous cyberattacks represent a specific subset of digital weapons in which algorithmic systems analyze large volumes of cyberthreat intelligence and make operational decisions without human intervention, a trend that has been observed in the growing use of machine-learning-based malware and dynamic command-and-control infrastructures capable of modifying their behavior in response to defensive countermeasures (S. Kaushik, 2023). As AI becomes embedded in offensive cyber capabilities, warfare increasingly reflects the logic of algorithmic speed, predictive analytics, and automated decision-making, characteristics that distinguish modern cyber conflict from earlier forms of digital intrusion.

AI warfare extends beyond traditional cyber operations by integrating artificial intelligence into broader military systems, including autonomous drones, decision-support platforms, and intelligent targeting systems. Scholars analyzing the evolution of military automation have emphasized that AI-driven systems can evaluate battlefield information with unprecedented speed, potentially supporting or replacing human decision-makers in areas such as threat detection, mission planning, and strategic assessment (Kumar & Mallipeddi, 2022). In the cyber domain, AI warfare involves the development of tools capable of autonomously identifying network vulnerabilities, generating novel exploits, and executing attacks across distributed digital environments. This integration of autonomy into offensive cyber operations marks a departure from earlier forms of cyber conflict, where human operators played a central role in planning and executing intrusions. The rising sophistication of algorithmic tools allows adversaries to conduct operations that are faster, more adaptive, and more difficult to detect, thereby shifting the balance of power between attackers and defenders.

The growing relevance of digital weapons in military doctrine reflects a recognition among states that cyber operations offer strategic advantages that traditional weapons cannot match. Scholars examining state practice have shown that cyber capabilities enable low-cost, deniable, and scalable means of achieving strategic objectives, whether through espionage, sabotage, information manipulation, or direct attacks against critical infrastructure (Gisel et al., 2020). Moreover, the integration of autonomous systems into national security strategies has accelerated as states perceive digital weapons as essential to maintaining technological superiority, deterring adversaries, and shaping geopolitical outcomes. The 2008 cyberattacks against Georgia, analyzed through the lens of just war doctrine, demonstrated how cyber operations can function as a precursor or complement to conventional military action (Yuliantiningsih, 2021). These developments underscore that digital weapons are no longer theoretical constructs but active components of modern strategic competition, influencing military planning, alliance behavior, and international diplomacy.

As the use of digital weapons expands, longstanding principles of International Humanitarian Law (IHL) face significant challenges. Scholars have observed that applying the principles of distinction, proportionality, and precaution to cyber operations is complicated by the interconnected nature of digital infrastructure, which often blurs the boundaries between civilian and military objects (Zuhra & Almira, 2021). The difficulty of distinguishing civilian networks from military systems becomes even more pronounced in autonomous cyberattacks, where self-learning algorithms may select targets based on statistical correlations rather than explicit legal criteria. Legal analyses highlight that attribution remains one of the most significant obstacles to enforcing IHL in the cyber domain, as the decentralized and anonymizable nature of digital operations makes it extremely difficult to identify perpetrators with sufficient certainty to trigger state responsibility or collective security mechanisms (Jiang, 2019). Furthermore, the inability of states to fully predict or control the behavior of autonomous systems raises questions about accountability, particularly when actions taken by AI do not reflect explicit human intent.

The expansion of cyber hostilities has prompted scholars and legal practitioners to call for a reconsideration of how IHL principles apply in digital contexts. Research examining the protection of civilians during cyber operations emphasizes that even non-kinetic digital attacks can have profound humanitarian consequences, such as disabling hospitals, water treatment facilities, or financial systems that are essential to civilian survival (Casey-Maslen & Mwale, 2021). This reality challenges traditional interpretations of "attacks" under IHL, which historically relied on the infliction of physical damage. Analyses of cyber operations during armed conflict have shown that digital effects can produce harm equivalent to or greater than traditional weapons, despite their non-physical nature, thereby necessitating an expansion of legal frameworks to accurately account for contemporary threats (Ali, 2022). Such concerns have led to increased attention to digital diplomacy and multilateral efforts aimed at clarifying legal norms governing cyber conflict, including proposals for extending IHL governance across new digital domains (Dumitru & Bodoni, 2022).

The evolution of digital weapons also raises ethical and operational concerns, particularly in relation to human decision-making and command responsibility. Scholars have highlighted the problem of "automation bias," wherein human operators may overly rely on algorithmic outputs during conflict, potentially leading to errors in judgment or unlawful targeting decisions (Kleemann, 2021). The delegation of critical functions to autonomous systems complicates the notion of meaningful human control, an emerging principle in debates surrounding autonomous weapons. Additionally, as states integrate cyber capabilities into broader defense architectures, there is a growing need for legal frameworks to address cross-border implications, escalation risks, and the possibility that autonomous cyberattacks might unintentionally trigger armed conflict by misinterpreting signals or misidentifying targets. These risks are compounded by public perceptions of cyber threats, which shape political and military responses, as demonstrated in recent analyses of how alliance commitments are influenced by public opinion during cyber crises (Gomez, 2023).

This article examines how digital weapons, autonomous cyberattacks, and AI warfare challenge existing legal and ethical frameworks, with particular emphasis on International Humanitarian Law. It employs a scientific narrative review method combined with descriptive analysis to synthesize insights from legal scholarship, cyber operations research, and emerging developments in AI-enabled conflict. The article aims to provide a comprehensive understanding of the regulatory gaps that hinder effective governance of digital weapons while proposing conceptual foundations for developing coherent legal frameworks. Because contemporary conflicts increasingly involve rapid, automated, and globally distributed forms of digital aggression, regulatory innovation has become essential to prevent uncontrolled escalation, reduce civilian harm, and ensure accountability across complex, multi-layered digital environments.

The urgency of regulating digital weapons is amplified by several factors. First, the unparalleled speed of autonomous cyber systems increases the likelihood of escalation, as attacks may unfold faster than human actors can interpret or respond. Second, attribution challenges hinder accountability and weaken deterrence, giving adversaries opportunities to exploit legal ambiguity and anonymity. Third, unpredictability in AI systems means that even well-intentioned actors may inadvertently trigger widespread harm if autonomous tools behave in unexpected ways. These combined risks illustrate why the development of robust regulatory frameworks is not merely desirable but necessary for maintaining international peace, protecting civilians, and ensuring that technological progress does not outpace the capacity of law to govern its consequences.

This introduction therefore sets the analytical foundation for exploring the legal, ethical, and strategic implications of digital weapons. By situating the emergence of AI-enabled cyber warfare within broader transformations in military doctrine and international relations, the discussion highlights the challenges that lie ahead for states, scholars, and international institutions as they attempt to reconcile evolving technologies with the enduring principles of humanitarian protection and the rule of law.

2. Autonomous Cyberattacks and AI Warfare: Technical and Operational Foundations

The rapid evolution of artificial intelligence and machine-learning techniques has fundamentally altered how cyber operations are conceived, implemented, and escalated in modern conflict environments. Autonomous cyberattacks now represent a growing category of digital aggression characterized by their ability to act independently of continuous human oversight, adapting to network conditions and optimizing their behavior during execution. Scholars analyzing the transformation of cyber operations note that AI-driven techniques increasingly allow offensive tools to make real-time decisions about reconnaissance, infiltration, and evasion, thereby amplifying their operational impact and strategic relevance (Tang et al., 2023). The incorporation of adaptive learning models enables digital weapons to evolve during a mission, adjusting to changing conditions in ways that traditional, manually controlled malware could not achieve. This shift marks a decisive movement toward algorithmically mediated warfare in which speed, autonomy, and unpredictability replace the slower, linear processes characteristic of earlier cyber operations.

Autonomous cyber systems exhibit several defining characteristics that differentiate them from conventional digital weapons. One of the most significant is self-propagation, the ability of a tool to independently scan, identify, and exploit vulnerabilities across networks without explicit human direction. Historical examples of fast-spreading malware demonstrate the disruptive potential of automated propagation, but AI-enabled systems extend this capability by leveraging pattern recognition and probabilistic modeling to determine optimal paths of infection (S. Kaushik, 2023). Self-selection of targets represents another critical feature, as machine-learning algorithms allow cyber tools to evaluate the type, configuration, and strategic importance of network assets, autonomously prioritizing targets based on predefined mission parameters or inferred operational value. The inclusion of adaptive learning further enhances system autonomy by enabling a cyber weapon to modify its behavior in response to detection indicators, environmental changes, or defensive countermeasures, a development noted in recent analyses of graph-based cyberthreat intelligence extraction (Tang et al., 2023). Such capabilities create a dynamic operational footprint that complicates both defensive responses and post-attack attribution.

Offensive AI capabilities extend beyond target selection and propagation, introducing new levels of sophistication in the design, execution, and concealment of cyberattacks. Research examining advanced malware techniques shows that machine learning can be used to optimize exploit strategies, refine evasion methods, and generate polymorphic variants that undermine signature-based detection systems (P. Kaushik, 2023). Autonomous penetration testing tools can conduct high-speed vulnerability assessments, prioritize attack vectors, and autonomously escalate privileges once inside a target environment. Intelligent deception represents another key application of AI in offensive cyber operations, as systems can analyze defensive tool outputs and craft misleading signals or decoy traffic to obscure attack pathways. The capacity for AI systems to integrate multiple offensive functions—scanning, infiltrating, learning, adapting, and concealing—creates a complex operational landscape in which cyberattacks evolve at machine speed, often outpacing human capacity to monitor or control them.

AI warfare extends beyond the cyber domain into hybrid operational environments where artificial intelligence interacts with physical military systems. Scholars examining military automation highlight that AI-driven target recognition systems can process imagery, sensor data, and geospatial information far more rapidly and accurately than human operators, enabling faster and potentially more lethal decision-making cycles (Kumar & Mallipeddi, 2022). Drone swarms represent one of the most visible examples of this convergence, as large numbers of autonomous aerial vehicles coordinate their movements, distribute computational tasks, and execute collective missions with minimal human oversight. AI-enhanced command-and-control systems similarly influence strategic planning and operational coordination, offering predictive analytics, battlefield simulations, and decision-support tools designed to optimize resource allocation and anticipate adversary behavior. These systems reflect a broader shift in warfare toward algorithmic coordination, where machine intelligence becomes embedded in multiple layers of military activity, from frontline engagement to high-level strategic command.

Understanding the operational foundations of AI warfare requires distinguishing between varying degrees of human involvement in autonomous systems. Human-in-the-loop systems require human approval for critical decisions, particularly those involving the use of force. Although these systems incorporate automation, they maintain human oversight to ensure compliance with legal and ethical standards. Human-on-the-loop systems allow machine processes to execute most functions autonomously, but a human operator supervises them and can intervene when necessary. Scholars observing the evolution of cyber operations note that these systems can still pose significant risks when operators become overly reliant on automated outputs or when the pace of operations exceeds human response times (Kleemann, 2021). Human-out-of-the-loop systems, by contrast, operate independently once activated, making decisions about timing, targeting, and execution without human intervention. In the cyber domain, such systems raise profound questions about accountability and control, as their actions may diverge from human intent or generate unanticipated consequences due to adaptive learning processes.

The escalation risks associated with autonomous cyberattacks and AI warfare arise from the combination of speed, unpredictability, and reliance on machine-driven decision-making. Analysts of cyber conflict have long warned that high-speed automated interactions between offensive and defensive systems can create rapid escalation pathways that outpace human comprehension, leading to unintended consequences or inadvertent conflict expansion (Schmitt, 2022). Unpredictability in machine-learning models exacerbates this risk, as algorithms trained on dynamic data may react differently to novel conditions, potentially misclassifying targets or misinterpreting defensive actions as threats. Automation bias further amplifies escalation risks, as human operators may defer to algorithmic recommendations even when those recommendations are flawed or legally problematic, a concern highlighted in research addressing the humanization of IHL in the context of cyber warfare (Kleemann, 2021). The loss of human control is particularly troubling in systems that evolve during operation, as adaptive models can generate behaviors that developers did not explicitly design or anticipate.

The relevance of autonomous cyberattacks extends to both state and non-state actors, reflecting the democratization of digital weapons and the accessibility of AI development tools. States continue to dominate high-end AI warfare capabilities, integrating autonomous tools into intelligence operations, cyber defense strategies, and offensive military doctrines. Studies examining cyber operations during armed conflict show that state actors increasingly rely on automated systems to support or conduct missions, leveraging digital tools to target critical infrastructure, conduct espionage, or prepare the battlefield for kinetic operations (Gisel et al., 2020). However, non-state actors—including terrorist organizations, cyber mercenaries, and advanced persistent threat (APT) groups—are also adopting AI-enabled tools as part of their operational portfolios. Analysts examining the legal frameworks for protecting civilians in cyber warfare note that the rise of non-state cyber actors increases the difficulty of applying IHL principles, as decentralized groups may exploit the anonymity afforded by digital operations and the lack of clear attribution mechanisms (Igakuboon, 2022). The expanding participation of non-state actors not only intensifies conflict dynamics but also challenges traditional models of deterrence and responsibility in cyberspace.

The involvement of non-state actors is further complicated by the strategic and operational advantages conferred by AI-driven tools. Research on cyberterrorism emphasizes that automated systems can enable smaller groups to execute complex operations that previously required substantial expertise or resources, thereby lowering the threshold for conducting disruptive attacks (Casey-Maslen & Mwale, 2021). Cyber mercenaries operating on behalf of states or private clients can deploy autonomous tools to infiltrate networks or disrupt services with limited direct oversight, making attribution even more complex. APT groups, often linked to state intelligence services, are increasingly incorporating machine-learning-based reconnaissance tools to enhance their stealth, persistence, and adaptability, enabling them to remain undetected inside networks for extended periods. The presence of these actors contributes to an environment where cyber operations are continuous, multi-layered, and increasingly resistant to defensive interventions.

The convergence of AI and cyber operations has also encouraged states to invest in digital diplomacy and strategic communication as part of broader AI warfare strategies. Analysts exploring the extension of IHL into information spaces highlight that states are leveraging digital technologies not only to conduct offensive operations but also to influence international norms, negotiate cyber agreements, and shape global perceptions of digital security (Dumitru & Bodoni, 2022). Public opinion plays an important role in shaping national responses to cyber incidents, as demonstrated in research analyzing how citizens perceive alliance commitments during cyber crises, which in turn affects political pressure and decision-making (Gomez, 2023). These dynamics illustrate that AI warfare is not confined to military or intelligence contexts but extends into diplomatic, social, and psychological domains, thereby broadening the scope of activities that autonomous systems influence.

As AI systems continue to evolve, the line between offensive and defensive applications becomes increasingly blurred. Researchers examining traditional cybercrime and penetration testing techniques observe that machine-learning-based tools can be repurposed or redirected across contexts, creating dual-use challenges that complicate regulatory efforts (P. Kaushik, 2023). The ability of algorithms to autonomously map networks, detect vulnerabilities, and generate optimized attack patterns means that the same tools used to strengthen cybersecurity can also be manipulated to launch sophisticated attacks. Supply chain systems, which rely on interconnected digital infrastructures, are particularly vulnerable to automated attacks, as disruptions in logistics, production, or resource allocation can cascade across global networks, a trend highlighted in research examining the impact of cybersecurity on supply chain management (Kumar & Mallipeddi, 2022). These vulnerabilities underscore the systemic risks posed by autonomous cyberattacks, especially when adversaries deploy tools capable of operating across multiple sectors simultaneously.

The technical and operational foundations of autonomous cyberattacks and AI warfare therefore reveal a complex landscape characterized by rapid evolution, increasing autonomy, and expanding participation from both state and non-state actors. The integration of AI into digital weapons has transformed the speed, scale, and sophistication of cyber operations while introducing profound risks related to control, attribution, and unintended escalation. Understanding these foundations is essential for developing legal and regulatory frameworks capable of addressing the challenges posed by autonomous digital conflict in an interconnected and rapidly changing global security environment.

3. Challenges for International Humanitarian Law (IHL) in Regulating Digital Weapons

The rapid integration of artificial intelligence, adaptive algorithms, and autonomous functionalities into digital weapons has exposed profound gaps in the existing legal architecture of International Humanitarian Law. Although IHL was designed to regulate the conduct of hostilities in diverse contexts, the emergence of AI-enabled cyber operations presents new forms of harm, new categories of actors, and new mechanisms of attack that do not align neatly with traditional legal categories. Scholars examining the intersection of cyber warfare and humanitarian protection have emphasized that IHL’s core principles—distinction, proportionality, necessity, and precaution—face unprecedented stress in digital environments where the boundaries between civilian and military systems are blurred, where the effects of cyber operations unfold in unpredictable ways, and where the attribution of harmful acts is often obscured by the layers of anonymity inherent in cyberspace (Gisel et al., 2020). The resulting doctrinal challenges necessitate critical analysis of how existing legal norms can be interpreted, adapted, or expanded to govern digital weapons effectively.

The principle of distinction, which requires belligerents to distinguish at all times between civilian objects and military objectives, becomes extraordinarily complex in the context of cyber operations. Civilian and military digital infrastructures are frequently intertwined, with critical systems—such as energy grids, health networks, financial platforms, and communication systems—coexisting on shared or overlapping digital architectures. Research evaluating IHL protections during cyber operations stresses that even a narrowly targeted operation may produce ripple effects across interconnected civilian systems, thereby challenging the requirement to avoid harming civilian objects (Igakuboon, 2022). Autonomous cyber tools compound this challenge by employing machine-learning models that prioritize efficiency and operational effectiveness rather than legal distinctions, potentially selecting targets based on algorithmic correlations rather than explicit legal significance. The principle of proportionality similarly becomes difficult to implement, as the indirect or delayed effects of cyber operations may be extremely difficult to predict, thereby complicating any assessment of whether anticipated harm to civilians is excessive relative to the military advantage pursued (Casey-Maslen & Mwale, 2021). Because autonomous systems may alter their behavior during execution, they can generate outcomes that diverge from the expectations of human operators, undermining attempts to evaluate proportionality at the time of planning.

The principle of necessity—which requires that the use of force be strictly necessary to achieve a legitimate military objective—also becomes less straightforward when cyber weapons are capable of acting autonomously, learning from their environment, or escalating without direct human intervention. Legal scholars have expressed concern that the speed and autonomy of AI-driven cyber operations may prompt states to deploy digital weapons more readily, treating them as low-cost alternatives to conventional force while underestimating their potential humanitarian impact (Ali, 2022). The precautionary principle, which mandates that belligerents take all feasible precautions to avoid or minimize harm to civilians, is similarly strained in environments where the behavior of digital weapons cannot be fully anticipated. As adaptive algorithms respond dynamically to defensive measures or environmental changes, the feasibility of meaningful precautions becomes uncertain, raising questions about whether states can meet their legal obligations when deploying systems characterized by unpredictability.

Attribution represents one of the most formidable doctrinal challenges in regulating autonomous cyberattacks under IHL. Scholars analyzing the legal implications of attribution emphasize that cyber operations can be routed through multiple intermediaries, anonymized through encryption, or executed by systems that evolve their own attack pathways, making it extremely difficult to identify the responsible actor with the degree of certainty required for establishing state responsibility (Jiang, 2019). Attribution becomes even more complex when autonomous systems modify their code or decision-making logic during deployment, raising questions about whether responsibility lies with the commander who authorized the operation, the developer who designed the algorithm, or the state that deployed the system. This diffusion of responsibility challenges core notions of command responsibility and state accountability that form the backbone of IHL enforcement. Some researchers argue that the anonymity inherent in cyberspace encourages states and non-state actors to engage more aggressively in cyber operations, exploiting the difficulty of attribution to avoid legal consequences (Zuhra & Almira, 2021). The rise of cyber mercenaries and advanced persistent threat groups further complicates the attribution landscape, as these actors often operate in legal gray zones, carrying out attacks that states may deny or disavow despite benefiting from their actions.

The problems of foreseeability and predictability of harm, central to the application of IHL, become particularly acute when dealing with adaptive AI systems. Adaptive algorithms do not simply execute predefined instructions; they evolve by interpreting data, identifying patterns, and adjusting strategies. Scholars analyzing offensive cyber tools note that such systems may generate novel behaviors during operation, producing consequences that developers or commanders could not have reasonably foreseen (Tang et al., 2023). When a cyber weapon autonomously alters its propagation strategy, selects new targets, or overrides human commands due to algorithmic optimization processes, traditional legal inquiries into intent and fault become difficult to apply. The unpredictability of adaptive systems challenges the principle of foreseeability that underlies the assessment of proportionality and the evaluation of unlawful attacks. IHL was constructed around the assumption that human actors make deliberate decisions about target selection and use of force. In contrast, autonomous systems can make operational choices that are opaque even to their creators, drawing into question how legal responsibility can be assigned when harm results from decisions made by an algorithmic entity.

The dual-use nature of digital infrastructure further complicates the application of IHL to AI-driven cyber operations. Digital networks such as cloud computing centers, telecommunications hubs, and internet exchange points serve both civilian and military functions, making them potential targets for digital weapons even though they are essential for civilian life. Analyses of cyber operations have demonstrated that attacks on digital infrastructure can disrupt hospitals, transportation networks, water treatment facilities, and other services critical to civilian survival, thereby producing humanitarian harm without causing physical destruction (Gisel et al., 2020). The interconnected nature of these systems means that even limited disruptions can have cascading effects, raising new questions about what constitutes a military objective in a highly digitized environment. Scholars examining the conduct of cyber warfare emphasize that IHL must confront the reality that digital environments lack the spatial separation that historically allowed belligerents to distinguish between civilian and military targets (Yuliantiningsih, 2021). This blurring of boundaries increases the risk of unlawful attacks and heightens the need for legal clarity in defining what constitutes a permissible target in cyber conflict.

Uncertainty also surrounds the threshold of "armed attack" within jus ad bellum when applied to AI-driven cyber operations. Traditional doctrines of self-defense assume a physical act of force that results in tangible destruction or injury. However, AI-enabled cyber operations can degrade systems, disrupt services, or destroy data without producing physical harm, raising questions about whether such actions meet the threshold required to justify the use of force in self-defense. Scholars evaluating cyber conflict argue that non-kinetic effects, such as disabling a nation's financial system or paralyzing critical infrastructure, may be equally or more harmful than conventional attacks, challenging the traditional interpretation of armed attack (Ali, 2022). The absence of consensus on whether digital harm constitutes an armed attack complicates the ability of states to respond lawfully to AI-driven operations, creating ambiguity that adversaries may exploit. This ambiguity becomes even more severe in scenarios where autonomous systems conduct operations at machine speed, potentially escalating a conflict before human decision-makers can assess the legality or proportionality of a response.

The regulation of pre-emptive, anticipatory, and continuous cyber operations introduces further complexity. Because cyber operations often unfold without clear temporal boundaries, it becomes difficult to determine when a conflict begins or ends, and when IHL obligations attach to state actions. Analysts examining the conduct of cyber hostilities argue that autonomous systems capable of continuous surveillance, penetration, or data manipulation blur the distinction between peacetime cyber activities and wartime operations (Protas & Joseph, 2020). Pre-emptive cyber operations, justified on the basis of preventing anticipated harm, raise legal concerns when adaptive algorithms generate threat assessments without human validation, potentially leading to premature or unjustified attacks. The legal framework governing anticipatory self-defense becomes strained in environments where predictive analytics and automated threat detection systems generate probabilistic assessments rather than clear evidence of imminent harm. Because autonomous systems may act continuously or intermittently without human oversight, it becomes difficult to impose temporal limits on hostilities, complicating the application of IHL rules governing the commencement, duration, and cessation of armed conflict.

One of the most debated questions in the regulation of digital weapons is whether data should be considered an object of attack under IHL. Traditional interpretations of "attack" rely on the concept of physical damage or destruction, yet many cyber operations target data directly by altering, deleting, encrypting, or corrupting information. Scholars analyzing the legal status of data destruction argue that non-physical harm can produce catastrophic effects on civilian populations, such as disabling medical records, disrupting transportation systems, or interfering with emergency services (Casey-Maslen & Mwale, 2021). Despite these consequences, IHL does not clearly define whether data constitutes a protected object, creating a gap that adversaries may exploit to launch harmful operations that technically avoid violating existing prohibitions. Some analyses suggest that the failure to categorize data as an object reflects the inadequacy of traditional legal frameworks in addressing contemporary technological realities (Gisel et al., 2020). The ambiguity surrounding the legal status of data highlights the need to modernize IHL definitions to reflect the importance of information systems in civilian life.

The Tallinn Manual is often cited as a reference for interpreting how existing international law applies to cyber operations, but its limitations are evident in the context of autonomous and AI-driven weapons. Although the Tallinn Manual provides detailed guidance on issues such as sovereignty, due diligence, and the conduct of hostilities, scholars emphasize that it remains a non-binding soft-law instrument developed through expert consensus rather than state practice (Schmitt, 2022). As a result, states may selectively adopt or ignore its principles, particularly when its interpretations conflict with national strategic interests. Moreover, the Manual does not fully address the unique challenges posed by adaptive AI systems, such as algorithmic unpredictability, autonomous target selection, or machine-learning-driven escalation pathways. Analysts evaluating the applicability of existing norms argue that the Tallinn Manual reflects a static understanding of cyber operations that does not account for the dynamic and evolving nature of AI-driven conflict (Dumitru & Bodoni, 2022). The lack of binding legal authority and the absence of comprehensive treatment of autonomous systems limit the Manual's usefulness as a regulatory tool for digital weapons.

These doctrinal challenges collectively demonstrate that IHL, while robust in its foundational principles, is ill-equipped to regulate the unique risks posed by autonomous cyberattacks and AI warfare. The conceptual gaps in attribution, foreseeability, target classification, and harm assessment underscore the need for renewed legal interpretation and updated frameworks that reflect contemporary realities. As digital weapons grow more sophisticated and autonomous, the limitations of existing IHL norms become increasingly evident, highlighting the urgency of adapting international law to govern emerging forms of digital conflict in a way that preserves humanitarian protections and promotes global stability.

4. Regulatory Approaches and Governance Models for AI and Cyber Warfare

Efforts to regulate AI-enabled cyber operations have evolved across multiple layers of international, regional, and national governance, yet the diversity of approaches reflects a fragmented regulatory landscape that struggles to keep pace with the rapid advancement of autonomous digital weapons. Existing frameworks attempt to extend traditional principles of international law into the cyber domain, but the unique characteristics of AI-driven cyber operations—speed, adaptiveness, opacity, and autonomy—challenge institutions designed for slower, more predictable forms of conflict. Scholars examining the interplay between cyber operations and humanitarian protection emphasize that current regulatory mechanisms remain limited in scope and inconsistent in implementation, leaving significant gaps in addressing the risks posed by autonomous escalation, self-learning algorithms, and algorithmically mediated military decision-making (Gisel et al., 2020). A coherent governance model for AI warfare remains elusive, with states adopting divergent interpretations of legal obligations and pursuing conflicting strategic interests in the digital domain.

The Tallinn Manual 2.0 represents one of the earliest comprehensive attempts to interpret how existing international law applies to cyber operations. Although its analyses cover issues including sovereignty, due diligence, and the conduct of hostilities, scholars note that its non-binding nature limits its practical influence, as states remain free to adopt or reject its interpretations based on strategic priorities (Schmitt, 2022). The Manual does not fully address the challenges of AI-enabled weapons, especially those involving self-learning systems or autonomous decision-making processes, reflecting its foundation in legal principles that predate contemporary advances in machine learning. Efforts within the United Nations framework provide a broader diplomatic context for cyber governance. The UN Group of Governmental Experts (GGE) has produced several consensus reports affirming that international law applies to cyberspace, yet the group has struggled to achieve progress due to geopolitical tensions, particularly among major powers with competing visions of digital sovereignty. Scholars analyzing state behavior in cyber diplomacy note that disagreements within the GGE often reflect broader geopolitical rivalries, limiting the group's ability to establish binding norms (Dumitru & Bodoni, 2022). The Open-Ended Working Group (OEWG), established to broaden participation beyond the limited membership of the GGE, has aimed to foster dialogue among a wider array of states, but like the GGE, it has faced difficulties in reaching consensus on key issues such as attribution, sovereignty, and acceptable peacetime behavior in cyberspace.

Regional governance initiatives provide additional layers of regulatory development. Within the European Union, the EU AI Act has emerged as one of the most ambitious attempts to regulate artificial intelligence across multiple domains. Although the Act focuses primarily on civilian applications, it includes defense exemptions that allow member states to deploy AI in military contexts under separate legal frameworks. The presence of these exemptions reflects the tension between safeguarding public welfare and maintaining flexibility for national security operations. Scholars analyzing the legal status of cyber operations argue that defense exemptions in regulatory instruments such as the EU AI Act create ambiguities that adversaries may exploit, particularly when dual-use technologies blur the boundaries between civilian and military applications (Kumar & Mallipeddi, 2022). Other regions, such as Southeast Asia and Eastern Europe, have developed national cyber strategies that emphasize digital sovereignty and resilience but differ significantly in their level of commitment to international norms, further complicating efforts to establish cohesive governance principles.

State positions on AI and cyber warfare vary widely, reflecting divergent strategic cultures, threat perceptions, and geopolitical ambitions. The United States has prioritized maintaining technological superiority while emphasizing the need for frameworks that preserve operational flexibility. Analysts examining U.S. doctrine suggest that the country views autonomy as essential for deterring adversaries and maintaining rapid response capabilities, yet remains cautious about fully delegating lethal decision-making to AI. China's approach emphasizes state control, cyber sovereignty, and the integration of AI into informationized warfare, reflecting its broader strategic goals in digital governance and military modernization (Tang et al., 2023). Russia, by contrast, emphasizes information control and hybrid warfare strategies, integrating cyber operations and AI tools into broader political and military campaigns aimed at undermining adversary cohesion. The European Union and the United Kingdom generally adopt more restrictive approaches, emphasizing human-centered AI principles and transparency in algorithmic processes, though their national defense strategies still pursue autonomous capabilities. Israel, known for its advanced cyber units and defense technologies, has embraced automation in intelligence and operational decision-making but has expressed support for meaningful human oversight to maintain compliance with international obligations. These differing state positions reveal fundamental disagreements about the appropriate degree of autonomy in military systems, undermining efforts to create unified global standards.

The responsibility of developers, private companies, and suppliers of dual-use technology is emerging as a significant governance challenge. Private-sector entities play a central role in creating machine-learning models, cyber defense tools, and digital infrastructure, often possessing capabilities that exceed those of states. Analysts examining civilian harm in cyber operations emphasize that private companies may inadvertently contribute to military cyber capabilities by developing dual-use technologies that can be weaponized or repurposed in conflict (Gisel et al., 2020). The growing involvement of technology contractors and cybersecurity firms further complicates the attribution of responsibility, as these actors operate in a commercial environment that does not always align with humanitarian or legal considerations. Scholars analyzing cyber operations note that the dual-use nature of AI tools means that regulatory frameworks must address the responsibilities of developers who enable both defensive and offensive applications (P. Kaushik, 2023). However, the absence of international standards for developer liability or algorithmic accountability leaves gaps that adversaries can exploit to justify harmful uses of AI-driven cyber capabilities.

Among the most significant regulatory gaps are those involving autonomous escalation, self-learning cyber weapons, and opaque algorithmic decision-making processes. Autonomous escalation occurs when digital systems respond to perceived threats without human intervention, potentially initiating or intensifying conflict. Analysts studying cyber hostilities highlight that machine-speed interactions between offensive and defensive algorithms can create rapid escalation pathways that exceed human ability to intervene effectively (Schmitt, 2022). Self-learning weapons introduce additional risks, as adaptive models may evolve beyond their initial design in ways that operators cannot fully control or predict. Opaque algorithms, particularly those employing deep learning, complicate legal review processes, as their decision-making logic may be difficult or impossible to interpret. Research examining offensive cyber tools illustrates that such opacity undermines attempts to evaluate compliance with IHL principles such as distinction and proportionality, which require clear understanding of how targeting decisions are made (Igakuboon, 2022). These gaps expose the inadequacy of traditional legal review mechanisms that rely on predictable, human-controlled weapons systems.

Proposals to strengthen governance frameworks often emphasize the importance of meaningful human control as a normative requirement for deploying autonomous systems. Scholars analyzing the humanization of war argue that human oversight is essential for ensuring compliance with humanitarian principles, particularly in contexts where autonomous systems may misinterpret data, misclassify targets, or escalate conflicts without clear authorization (Kleemann, 2021). Meaningful human control involves ensuring human supervision at critical decision points, including target selection, activation of cyber weapons, and assessment of potential harm. Other proposals call for algorithmic accountability in military systems, requiring states to document how AI models are trained, validated, and reviewed prior to deployment. Analysts studying cyber diplomacy argue that such accountability mechanisms are essential for establishing trust among states and reducing the risk of miscalculation or unintended escalation (Dumitru & Bodoni, 2022). Auditability and explainability are similarly important for regulating autonomous weapons, as they enable investigators to trace system behavior, identify causes of failure, and determine responsibility for unlawful acts.

International oversight mechanisms represent another potential avenue for governance. Some scholars propose adapting arms-control models from nuclear and chemical weapons regimes, leveraging verification mechanisms, transparency measures, and confidence-building processes to manage AI warfare risks. The nuclear non-proliferation regime provides a model for intrusive inspections and reporting requirements, though AI technologies differ significantly from nuclear capabilities in terms of ease of proliferation and dual-use potential. The Chemical Weapons Convention offers a precedent for banning entire categories of weapons, though applying such a model to AI-driven cyber weapons would require unprecedented levels of international agreement. Analysts examining public perceptions of cybersecurity note that transparency frameworks and confidence-building measures could reduce the risks of misinterpretation during cyber crises, particularly when attribution remains uncertain (Gomez, 2023). However, confidence-building measures require political willingness that may be lacking in a geopolitical environment characterized by rapid technological competition and strategic mistrust.

The feasibility of an international treaty banning fully autonomous cyber weapons remains a topic of considerable debate. Some scholars argue that a ban is necessary to prevent destabilizing arms races and reduce the risk of harm to civilians, emphasizing that autonomous systems pose unpredictable and potentially catastrophic risks that justify preemptive prohibition (Casey-Maslen & Mwale, 2021). Others contend that the dual-use nature of AI technologies, the difficulty of verification, and the strategic incentives for states to maintain autonomy in digital warfare make a comprehensive ban impractical. Research examining state attitudes toward cyber regulation suggests that many states are reluctant to accept constraints that might limit their offensive capabilities, particularly when rivals are unlikely to reciprocate (Protas & Joseph, 2020). The absence of shared definitions for autonomous cyber weapons further complicates treaty negotiations, as states disagree on whether autonomy should be defined in terms of target selection, mission execution, or machine-learning capability. While a treaty banning fully autonomous cyber weapons would contribute significantly to global stability, its implementation would require unprecedented cooperation and verification mechanisms that currently appear unlikely given geopolitical divisions.

Despite these challenges, regulatory approaches and governance models continue to evolve as states, international organizations, and private entities recognize the need to address the risks associated with AI-enabled cyber warfare. A comprehensive governance model will likely require a multilayered approach that integrates legal norms, technical standards, oversight mechanisms, and international cooperation. By synthesizing insights from existing frameworks and acknowledging current limitations, it becomes clear that innovative regulatory solutions are essential for ensuring that technological advancements in cyber warfare do not undermine global stability or human security.

5. Toward a Coherent Legal and Ethical Framework for Digital Weapons

The emergence of autonomous cyberattacks and AI-enabled military systems has created a landscape in which traditional legal norms and security frameworks must adapt to technologies capable of operating at machine speed, learning autonomously, and interacting across global digital environments. Developing a coherent legal and ethical framework for digital weapons requires integrating insights from technology studies, security research, and International Humanitarian Law to produce principles capable of guiding state behavior while preserving human-centered values. Scholars examining the protection of civilians in cyber conflict emphasize that this integration must confront the reality that autonomous systems introduce new forms of unpredictability, obscure lines of accountability, and expand the scope of harm that war can produce in digital and physical domains (Gisel et al., 2020). As AI-driven systems become more deeply embedded in offensive and defensive cyber operations, normative approaches must be redesigned to address risks that were previously inconceivable within traditional frameworks of armed conflict.

One of the central elements of a future regulatory architecture is the incorporation of meaningful human control into all phases of autonomous cyber operations. Researchers studying the humanization of war argue that human oversight is essential for ensuring compliance with humanitarian principles and preventing algorithmic systems from making decisions that violate ethical or legal standards (Kleemann, 2021). Meaningful human control requires that operators retain the ability to understand, supervise, and intervene in the operation of digital weapons, especially when those systems are capable of evolving beyond their initial programming. Ensuring such control becomes more complex as cyber weapons integrate machine-learning mechanisms that allow for adaptation and independent decision-making. Analysts examining offensive cyber capabilities note that autonomous systems can initiate or escalate actions without explicit authorization, highlighting the need for layered safeguards that maintain human judgment as the ultimate arbiter of critical decisions (Schmitt, 2022). Establishing technological, organizational, and legal standards for human oversight will be essential for preserving ethical decision-making in environments dominated by machine-speed conflict.

Defining accountability represents another foundational requirement for a coherent legal framework. Current doctrines of responsibility are challenged by the opacity, adaptiveness, and unpredictability of AI systems used in cyber warfare. Scholars examining the challenges of attribution in cyber operations emphasize that actions carried out by autonomous tools may involve contributions from software developers, data scientists, commanders, and states, creating a dispersion of agency that complicates traditional accountability frameworks (Jiang, 2019). When an autonomous system misidentifies a target or causes unintended harm, determining who bears legal responsibility becomes a complex endeavor requiring forensic insight into algorithmic decision paths. Analysts studying offensive AI capabilities note that uncertainty surrounding system behavior undermines the foreseeability required by IHL to assign liability for unlawful acts (Tang et al., 2023). A future governance model must therefore establish clear standards for documenting system design, training data, operational parameters, and human oversight structures to enable meaningful accountability when harm occurs. Without robust accountability mechanisms, states may be incentivized to rely on autonomous tools as a means of obscuring responsibility for prohibited actions.

A coherent legal framework must also include shared definitions of autonomous cyber weapons, a prerequisite for meaningful international regulation. At present, states disagree on how to define autonomy in digital systems, with some focusing on the ability to select targets independently, others emphasizing self-learning capabilities, and still others framing autonomy in terms of operational independence from human oversight. Scholars analyzing international cyber law emphasize that the absence of shared definitions hampers efforts to establish norms, negotiate treaties, or implement confidence-building measures (Dumitru & Bodoni, 2022). Without consensus, states may adopt incompatible regulatory measures or exploit definitional ambiguity to justify the development of increasingly autonomous systems. Establishing shared terminology will require multidisciplinary collaboration across technical, legal, and diplomatic communities to ensure that definitions capture the complexities of AI-driven systems while remaining grounded in operational reality.

Strengthening due-diligence obligations among states represents another essential component of a future governance model. Traditional notions of due diligence require states to ensure that their territory is not used to harm other states, but applying this principle to cyber operations presents unique challenges. Scholars studying cyber espionage and digital sabotage argue that states must implement measures to prevent non-state actors from deploying harmful AI tools from within their borders, particularly given the accessibility of machine-learning models and dual-use technologies (Igakuboon, 2022). Due diligence in the AI era should include requirements for monitoring domestic cyber activity, regulating AI development practices, and cooperating with other states in identifying harmful operations. Analysts examining cyber conflict note that the dual-use nature of digital infrastructure complicates efforts to prevent misuse, as many civilian systems can be repurposed for offensive cyber activities (Kumar & Mallipeddi, 2022). A strengthened due-diligence framework would require states to adopt policies aimed at reducing vulnerabilities in their critical infrastructure, promoting responsible AI innovation, and sharing information about emerging threats.

Ethical considerations must be integrated into any legal framework, as the humanitarian consequences of digital weapons extend beyond traditional battlefield concerns. Scholars examining cyber terrorism and civilian protection highlight that cyber operations can disrupt essential services such as health care, water supply, and emergency response systems, producing humanitarian harm even without physical destruction (Casey-Maslen & Mwale, 2021). Protecting critical infrastructure therefore becomes an ethical imperative, requiring that states refrain from targeting systems essential to civilian survival. Fairness and transparency in AI targeting also represent key ethical concerns, as opaque algorithms raise the risk of discriminatory or erroneous outcomes. Analysts studying algorithmic deception and malware optimization argue that as AI systems become more complex, their decision-making processes become less interpretable, increasing the risk of harm to civilians due to misclassification or bias embedded in training data (P. Kaushik, 2023). Integrating ethical principles into AI design processes—such as by adopting value-sensitive design, algorithmic transparency, and risk assessments—can help ensure that digital weapons adhere to humanitarian norms even in high-speed operational environments.

Research gaps remain significant in areas that are critical for establishing a robust governance framework. System verification represents one of the most important gaps, as there is currently no widely accepted method for evaluating whether autonomous cyber systems behave as intended under operational conditions. Scholars studying AI-assisted malware detection emphasize that system verification is particularly challenging because adaptive models may evolve during use, producing behaviors that were not visible during testing (Tang et al., 2023). AI explainability represents another gap, as the ability to interpret machine decision processes is essential for legal review, accountability, and post-incident investigation. The lack of explainability becomes especially problematic in high-speed combat environments where autonomous systems make decisions at machine speed, leaving little opportunity for human oversight. Analysts examining the integration of AI in military systems note that without explainability, it becomes nearly impossible to determine whether system behavior was lawful or consistent with operational objectives (Kleemann, 2021).

Cross-border governance presents an additional challenge, as digital operations routinely transcend national boundaries, complicating jurisdictional questions and the enforcement of international norms. Scholars analyzing digital diplomacy highlight that states must develop new mechanisms for cooperation, information sharing, and attribution to manage cross-border AI threats effectively (Dumitru & Bodoni, 2022). Non-state actors, including cyber mercenaries and terrorist organizations, further complicate governance, as these groups are not bound by traditional norms and may exploit autonomous systems to conduct disruptive operations. Research examining the rise of such actors emphasizes the need for cyber norms that explicitly address non-state behavior and establish expectations for state responsibility in preventing harmful activities originating from their territory (Igakuboon, 2022). Developing norms for non-state actors will be essential for preventing the proliferation of autonomous cyber weapons and ensuring that states do not outsource responsibility to private entities or proxy groups.

The development of a coherent legal and ethical framework for digital weapons requires sustained interdisciplinary collaboration, innovative governance models, and a commitment to preserving humanitarian principles in an era of rapid technological transformation. By integrating meaningful human control, clarifying accountability structures, establishing shared definitions, strengthening due diligence, and addressing ethical considerations, such a framework can help guide the responsible use of AI in warfare. Filling existing research gaps related to verification, explainability, cross-border governance, and non-state actors will further strengthen the foundation for future regulation. Through these combined efforts, the international community can work toward a governance model that mitigates risks, preserves global stability, and ensures that digital weapons evolve in a manner consistent with legal and ethical imperatives.

6. Conclusion

The rapid evolution of autonomous cyberattacks and AI-enabled warfare has fundamentally reshaped the strategic, ethical, and legal landscape of modern conflict. Digital weapons that operate at machine speed, adapt in real time, and make targeting decisions independently of human operators mark a departure from earlier forms of cyber activity that were more predictable and easier to regulate. As states increasingly integrate artificial intelligence into both offensive and defensive cyber capabilities, the need for a coherent regulatory framework becomes not simply an academic concern but a pressing necessity for international peace and security. The risks posed by autonomous systems—ranging from unintended escalation to large-scale civilian harm—illustrate how profoundly digital warfare challenges the assumptions upon which traditional laws of armed conflict were built.

Regulating these technologies requires acknowledging and addressing several unique features that distinguish AI-driven cyber operations from conventional weapons. Autonomous systems introduce unprecedented speed and operational autonomy, eroding the time available for human judgment and potentially amplifying the consequences of error. Adaptive algorithms complicate predictions about weapon behavior, as their decision-making evolves dynamically during deployment. The interconnectedness of global digital infrastructure means that attacks directed at military assets can produce cascading effects across civilian networks, disrupting essential services and endangering populations far removed from the intended target. These characteristics challenge foundational principles of International Humanitarian Law and expose gaps that existing doctrines are not yet equipped to fill.

At the same time, states differ significantly in their approaches to regulating AI warfare. Some prioritize operational flexibility and technological superiority, while others emphasize transparency, human oversight, and risk mitigation. These divergent positions complicate international negotiations and hinder efforts to craft unified norms or binding agreements. The involvement of private companies and dual-use technology developers adds further complexity, as these entities produce the tools that shape modern cyber capabilities yet operate outside traditional military structures. The increasing participation of non-state actors—including cyber mercenaries and technologically sophisticated terrorist groups—underscores the need for governance models that extend beyond state-centric frameworks.

Despite these challenges, several guiding principles can help build a coherent system of regulation. Integrating meaningful human control into the design, deployment, and oversight of autonomous systems is critical for ensuring that human judgment remains central to decisions involving harm and conflict escalation. Establishing clear accountability mechanisms can prevent the diffusion of responsibility that often accompanies the use of autonomous tools. Developing shared definitions of autonomous cyber weapons will lay the groundwork for treaties, verification mechanisms, and transparency measures. Strengthening due-diligence obligations can help states mitigate risks stemming from the misuse of dual-use technologies and prevent harmful operations emanating from their territory. Ethical considerations—including the protection of civilians, respect for critical infrastructure, and fairness in algorithmic decision-making—must serve as guiding values throughout these efforts.

Moving forward, researchers and policymakers face several urgent priorities. One is advancing verification methods capable of evaluating the behavior of autonomous systems under real-world conditions. Another is enhancing explainability in AI models so that operators, investigators, and legal authorities can interpret and assess system decisions. Cross-border governance will also be essential, as digital operations routinely transcend national boundaries and impose shared risks that no state can manage alone. Furthermore, establishing cyber norms for non-state actors will be necessary to address the growing influence of groups that operate beyond the reach of traditional diplomatic or legal tools.

Ultimately, the task of regulating digital weapons calls for a combination of legal innovation, technical expertise, and ethical commitment. International cooperation will be crucial, but it must be grounded in realistic assessments of geopolitical constraints and technological realities. While complete consensus among states may be difficult to achieve, incremental progress through confidence-building measures, transparency mechanisms, and multilateral agreements can contribute significantly to reducing risks. The long-term goal should be a governance system that preserves humanitarian protections, supports global stability, and ensures that technological advancements do not outpace society's ability to manage their consequences responsibly.

The future of conflict will be shaped in part by how the world chooses to govern the new domain of AI-enabled digital warfare. By addressing existing gaps, promoting responsible innovation, and reinforcing the centrality of human judgment, the international community can work toward a framework that balances military necessity with ethical imperatives. The development of such a framework will determine whether digital weapons become tools that destabilize the global order or instruments that can be effectively regulated within the bounds of international law.

Ethical Considerations

All procedures performed in this study were under the ethical standards.

Acknowledgments

Authors thank all participants who participate in this study.

Conflict of Interest

The authors report no conflict of interest.

Funding/Financial Support

According to the authors, this article has no financial support.

References

Ali, S. (2022). Legal Framework of Right of Self Defense in Cyber Warfare: Application Through Laws of Armed Conflict. *Journal of Development and Social Sciences*, *3*(II). https://doi.org/10.47205/jdss.2022(3-ii)96

Casey-Maslen, S., & Mwale, B. (2021). The Prohibition of Cyberterrorism as a Method of Warfare in International Law. *South African Yearbook of International Law*, *44*. https://doi.org/10.25159/2521-2583/7977

Dumitru, D., & Bodoni, C. (2022). Extension of International Humanitarian Law Order in the Information Area Through Digital Diplomacy. *Strategic Impact*, *80*(3), 86-102. https://doi.org/10.53477/1841-5784-21-18

Gisel, L., Rodenhäuser, T., & Dörmann, K. (2020). Twenty Years On: International Humanitarian Law and the Protection of Civilians Against the Effects of Cyber Operations During Armed Conflicts. *International Review of the Red Cross*, *102*(913), 287-334. https://doi.org/10.1017/s1816383120000387

Gomez, M. A. (2023). Public Opinion and Alliance Commitments in Cybersecurity: An Attack Against All? https://doi.org/10.31235/osf.io/bcwhu

Igakuboon, A. N. (2022). An Appraisal of the Legal Framework for the Protection of Civilians in Cyber-Warfare Under International Humanitarian Law. *International Journal of Research and Scientific Innovation*, *09*(07), 14-26. https://doi.org/10.51244/ijrsi.2022.9702

Jiang, Z. (2019). Regulating the Use and Conduct of Cyber Operations Through International Law: Challenges and Fact-Finding Body Proposal. *Lselr*, *5*, 59-88. https://doi.org/10.61315/lselr.42

Kaushik, P. (2023). Deep Learning Multi-Agent Model for Phishing Cyber-Attack Detection. *International Journal on Recent and Innovation Trends in Computing and Communication*, *11*(9s), 680-686. https://doi.org/10.17762/ijritcc.v11i9s.7674

Kaushik, S. (2023). Analysis of Blockchain Security: Classic Attacks, Cybercrime and Penetration Testing. https://doi.org/10.1109/mobisecserv58080.2023.10329210

Kleemann, S. (2021). Cyber Warfare and the "Humanization" of International Humanitarian Law. *International Journal of Cyber Warfare and Terrorism*, *11*(2), 1-11. https://doi.org/10.4018/ijcwt.2021040101

Kumar, S., & Mallipeddi, R. R. (2022). Impact of cybersecurity on operations and supply chain management: Emerging trends and future research directions. *Production and Operations Management*, *31*(12), 4488-4500. https://doi.org/10.1111/poms.13859

Protas, P., & Joseph, L. C. (2020). The Law of Armed Conflict in the Era of Cyber Technology: Assessing the Legal Challenges and Response in Tanzania. *Eastern Africa Law Review*, *47*(1), 95-139. https://doi.org/10.56279/ealr.v47i1.4

Schmitt, M. N. (2022). International Humanitarian Law and the Conduct of Cyber Hostilities: Quo Vadis? *Journal of International Humanitarian Legal Studies*, *13*(2), 189-221. https://doi.org/10.1163/18781527-bja10059

Tang, B., Wang, J., Qiu, H., Yu, J., Yu, Z., & Liu, S. (2023). Attack Behavior Extraction Based on Heterogeneous Cyberthreat Intelligence and Graph Convolutional Networks. *Computers Materials & Continua*, *74*(1), 235-252. https://doi.org/10.32604/cmc.2023.029135

Yuliantiningsih, A. (2021). Analisis Doktrin Perang Yang Adil (Just War) Dalam Kasus Serangan Siber Rusia Terhadap Georgia Tahun 2008. *Kosmik Hukum*, *21*(3), 175. https://doi.org/10.30595/kosmikhukum.v21i3.10613

Zuhra, A., & Almira, L. (2021). The LIMITATION OF CYBER WARFARE UNDER HUMANITARIAN LAW (Pembatasan Perang Siber Dalam Hukum Humaniter). *Teras Law Review Jurnal Hukum Humaniter Dan Ham*, *3*(1), 1-10. https://doi.org/10.25105/teras-lrev.v3i1.10741