

Autonomous Weapons and Cyber Warfare: Ethical Boundaries, International Law, and the Future of Human Security

Parhlad Singh Ahluwalia, Academician, Dr. Bhimrao Ambedkar Law University, Jaipur, Rajasthan, India

Abstract

The convergence of artificial intelligence, robotics, and offensive cyber capability has produced two of the most consequential developments in the modern conduct of armed conflict: autonomous weapons systems (AWS) capable of selecting and engaging targets with reduced or absent human intervention, and cyber warfare capable of disabling critical infrastructure, financial systems, and military command networks without a single soldier crossing a border. This paper undertakes an interdisciplinary review of the ethical, legal, and human-security dimensions of both phenomena, treating them as related manifestations of a broader shift toward the delegation of lethal and coercive decision-making to machines and code. Drawing on United Nations documentation, the Convention on Certain Conventional Weapons (CCW) process, the Tallinn Manual 2.0, International Committee of the Red Cross (ICRC) policy positions, and peer-reviewed scholarship, the paper synthesises the principal ethical objections to autonomous lethality — the accountability gap, the erosion of human dignity, and the risk of automation bias — alongside the doctrinal uncertainties that cyber operations pose for international humanitarian law (IHL) and the *jus ad bellum*. Original comparative tables and data visualisations are used to map state positions at the United Nations, document milestone incidents (Stuxnet, NotPetya, the Kargu-2 strike in Libya, and the Nagorno-Karabakh and Ukraine conflicts), and estimate the economic scale of cyber harm. The paper concludes that neither existing IHL nor the fragmented cyber-norms regime is adequate to the emerging threat environment, and that human security — rather than state security alone — should anchor the design of a future governance architecture. A treaty combining a prohibition on autonomous weapons that cannot be operated under meaningful human control with binding, attributable norms for state-sponsored cyber operations is proposed as the most coherent path forward.

Keywords: *autonomous weapons systems; lethal autonomous weapons; cyber warfare; international humanitarian law; meaningful human control; human security; Tallinn Manual; killer robots; artificial intelligence ethics; laws of armed conflict.*

1. Introduction

The twenty-first-century battlefield is no longer defined solely by territory, troop strength, or firepower. It is increasingly defined by algorithms, sensors, and code. Two developments in particular have reshaped the character of contemporary and future armed conflict: the emergence of autonomous weapons systems (AWS) capable of independently selecting and engaging targets, and the maturation

of cyber warfare as a domain of state and non-state confrontation that operates below, alongside, and sometimes instead of kinetic force [1,2]. Both developments share a common thread: they displace human judgement, whether partially or entirely, from decisions that traditionally required a human being to perceive, deliberate, and act. This displacement raises urgent ethical, legal, and security questions that existing frameworks of international humanitarian law (IHL), arms control, and human security were not designed to answer.

Autonomous weapons systems are commonly defined as weapons that, once activated, can select and engage targets without further intervention by a human operator [3]. The March 2020 Libyan civil war is widely cited as the first documented instance in which such a system — the Turkish-manufactured STM Kargu-2 loitering munition — may have engaged human targets in a "fire, forget and find" mode that did not require a live data link to an operator, according to the report of the United Nations Panel of Experts on Libya [4,5]. Whether or not this specific engagement constituted the first lethal use of a fully autonomous weapon remains contested among specialists, but the episode crystallised a shift that had been building for more than a decade: warfare was moving from remotely piloted systems toward systems that could, in principle, complete the "find-fix-track-target-engage-assess" cycle without a human in the loop [6].

Cyber warfare has followed a parallel trajectory. What began with relatively unsophisticated denial-of-service campaigns — such as the 2007 attacks against Estonian government, banking, and media infrastructure — evolved within a decade into operations capable of causing physical destruction, as demonstrated by the Stuxnet worm's sabotage of Iranian uranium-enrichment centrifuges at Natanz in 2010 [7,8], and into operations capable of inflicting billions of dollars in cascading economic damage across multiple countries simultaneously, as demonstrated by the 2017 NotPetya attack, which caused an estimated US\$10 billion in global losses despite masquerading as ransomware [9,10]. By 2025, the global cost of cybercrime and cyber-enabled disruption — encompassing but not limited to state-sponsored activity — was projected to reach US\$10.5 trillion annually, a figure that, if it were a national economy, would rank third in the world after the United States and China [11].

These two domains — physically embodied autonomous weapons and disembodied cyber operations — are frequently studied in isolation. Yet they are converging in practice. Autonomous weapons increasingly depend on networked sensor fusion, machine-learning-based target recognition, and software supply chains that are themselves vulnerable to cyber compromise; a hacked autonomous weapon is a qualitatively new threat that neither the law of armed conflict nor cyber-norms literature adequately anticipates [12]. Conversely, cyber operations increasingly incorporate autonomous or semi-

autonomous elements — self-propagating malware, AI-assisted vulnerability discovery, and automated intrusion-response systems — that reduce human oversight over offensive cyber action in ways structurally analogous to the "meaningful human control" debate in the kinetic domain [13]. This paper therefore treats autonomous weapons and cyber warfare as two expressions of a single underlying transformation: the delegation of life-and-death and infrastructure-critical decisions to non-human agents operating at machine speed.

The purpose of this paper is threefold. First, it maps the ethical terrain of this transformation, engaging with the principal philosophical objections to autonomous lethality — the "responsibility gap," the argument from human dignity, and the empirical literature on automation bias — and with the distinct but related ethical questions raised by cyber operations, including proportionality in a domain with poorly bounded effects and the ethics of anonymity and misattribution. Second, it examines the adequacy of the existing international legal architecture, principally IHL as codified in the 1949 Geneva Conventions and their 1977 Additional Protocols, the Convention on Certain Conventional Weapons (CCW) process on lethal autonomous weapons systems (LAWS), and the Tallinn Manual 2.0's restatement of how international law applies to cyber operations. Third, it situates both phenomena within the broader concept of human security — as articulated in the United Nations Development Programme's 1994 Human Development Report — arguing that a state-centric security paradigm is insufficient to capture the risks that autonomous weapons and cyber warfare pose to individuals, communities, and civilian infrastructure [14,15].

The paper proceeds as follows. Section 2 reviews the relevant literature across international law, military ethics, computer science, and security studies. Section 3 outlines the methodology and sources used. Section 4 characterises the technology and typology of autonomous weapons systems. Section 5 does the same for cyber warfare, including a comparative table of milestone incidents. Section 6 analyses the ethical boundaries at stake. Section 7 examines the international legal framework in detail. Section 8 presents case studies with an original comparative table. Section 9 surveys divergent stakeholder perspectives — major military powers, middle powers, civil society, and international organisations — in a structured comparison. Section 10 discusses implications for human security. Section 11 offers policy recommendations, and Section 12 concludes.

2. Literature Review

2.1 Autonomous Weapons Systems: Definitional and Ethical Scholarship

Scholarly engagement with autonomous weapons predates their battlefield deployment by more than a decade. Early and highly influential work by Sharkey argued that no existing or foreseeable machine-vision system could reliably discriminate between combatants and civilians under the conditions required by the IHL principle of distinction, making fully autonomous targeting inherently unlawful under contemporary technological conditions [16]. Asaro extended this critique into the domain of moral philosophy, arguing that the delegation of a lethal decision to a machine is a violation of human dignity independent of whether the machine's targeting decisions are statistically as accurate as, or more accurate than, a human soldier's, because the victim of lethal force has a moral claim to be killed (or spared) by a decision-maker capable of empathy and moral judgement [17]. This "dignity" argument has become one of the two dominant ethical objections in the literature, alongside the "responsibility gap" thesis advanced by Sparrow, who contends that autonomous weapons create a scenario in which no party — not the commander, not the programmer, not the machine itself — can be held morally or legally responsible for a wrongful death, producing an accountability vacuum inconsistent with the requirements of *jus in bello* [18].

Human Rights Watch and Harvard Law School's International Human Rights Clinic produced the report "Losing Humanity" in 2012, which was instrumental in launching the Campaign to Stop Killer Robots and in framing fully autonomous weapons as a matter requiring pre-emptive prohibition rather than post-hoc regulation [19]. The International Committee of the Red Cross has since taken an increasingly firm position, calling in 2021 for new legally binding rules that would prohibit unpredictable autonomous weapons outright and require meaningful human control over all others, particularly those directed against human targets [20]. Scharre, writing from a national-security perspective, offers a more permissive account, distinguishing between "narrow" autonomy in defensive systems (such as the Phalanx close-in weapons system, in use since the 1980s) and the far more ethically fraught prospect of offensive systems that autonomously select human targets, arguing that a blanket ban is neither technologically nor strategically realistic and that the more productive path is to codify standards of "appropriate levels of human judgment" [21]. This tension between prohibitionist and regulatory schools of thought runs throughout the literature and, as this paper shows, throughout the diplomatic record at the United Nations.

2.2 Cyber Warfare: Legal and Strategic Scholarship

The foundational legal text in this domain is the Tallinn Manual, produced under the auspices of the NATO Cooperative Cyber Defence Centre of Excellence and led by Michael N. Schmitt. The first edition (2013) addressed the law applicable to cyber operations that rise to the level of armed conflict; the expanded second edition, Tallinn Manual 2.0 (2017), broadened the analysis to peacetime legal regimes, articulating 154 "black-letter" rules covering sovereignty, state responsibility, human rights, and the law of the sea, air, and space as applied to cyber operations [22,23,24]. The Manual is explicitly non-binding and represents the opinions of its expert authors rather than state practice, but it has become the most widely cited reference point for how existing international law extends into cyberspace [25].

Rid offers an influential, if contested, argument that cyber "war" as traditionally conceived may be a category error: because cyber operations to date have rarely caused physical destruction or loss of life independently, Rid contends that most cyber incidents are better classified as sabotage, espionage, or subversion rather than acts of war in the Clausewitzian sense [26]. Lin and Zegart, by contrast, argue that this distinction is analytically useful but practically fragile, since the cumulative and cascading economic and infrastructural effects of major incidents — NotPetya being the paradigmatic case — can approach or exceed the harm caused by limited kinetic strikes, meaning legal categories built around physical destruction risk under-regulating a rapidly escalating threat [27]. The attribution problem is a second recurring theme: Rid and Buchanan demonstrate that reliable technical attribution of a cyber operation to a specific state actor is possible but resource-intensive, time-consuming, and inherently probabilistic, which complicates the invocation of state-responsibility doctrines that presume clear authorship [28].

2.3 Human Security as an Analytical Frame

The human security paradigm, introduced in the United Nations Development Programme's 1994 Human Development Report, redefined security away from the protection of state territory and sovereignty and toward "freedom from fear" and "freedom from want" at the level of the individual [14]. Kaldor's work on "new wars" complements this framework by documenting how contemporary conflicts increasingly blur the line between combatants and civilians, between public and private violence, and between war and organised crime — a blurring that both autonomous weapons (through their proliferation to non-state and criminal actors) and cyber warfare (through the routine targeting of civilian infrastructure such as hospitals, power grids, and financial systems) intensify [29]. Comparatively little scholarship, however, applies the human security frame jointly to autonomous

weapons and cyber warfare; most legal and ethical analysis remains siloed by domain. This paper responds to that gap by treating the two as a joint governance problem.

2.4 State Doctrine and Policy Literature

A fourth strand of literature examines the declared doctrine and procurement priorities of the states most active in developing autonomous and cyber capability. US policy, formalised in Department of Defense Directive 3000.09, does not prohibit autonomous weapons but requires that "appropriate levels of human judgment" be exercised over the use of force, alongside a rigorous senior-level review process for systems intended to operate with reduced human oversight [3]. Boulanin and Verbruggen, examining the Article 36 legal-review process required under Additional Protocol I for all new weapons, note that no state has yet published a completed Article 36 review of a fully autonomous weapons system, leaving the practical operation of this pre-existing legal safeguard largely untested for the technology it is arguably most needed to constrain [14]. Chinese policy statements at the CCW have called for a ban on the "use" of fully autonomous weapons while simultaneously supporting continued research, development, and production — a position Human Rights Watch characterises as internally inconsistent given China's own advanced AWS development programmes [1]. Russian representatives at the CCW have consistently opposed any new legally binding instrument, arguing that existing IHL is sufficient and that a treaty would be premature given the absence of an agreed technical definition of autonomy [44,45]. This divergence between declared doctrine and demonstrated capability recurs as a theme throughout the present paper's comparative analysis in Sections 7 and 9.

2.5 Summary of the Gap Addressed by This Paper

Three gaps recur across the literature reviewed above. First, ethical and legal analysis of autonomous weapons is rarely integrated with the parallel literature on cyber warfare, despite the two domains sharing structural features — reduced human control, distributed responsibility, and asymmetric proliferation risk. Second, most empirical case-study work focuses on single incidents (Stuxnet; the Libya Kargu-2 episode) rather than building comparative datasets that allow cross-case analysis of trends over time. Third, relatively few studies translate the human security paradigm into concrete governance recommendations that span both domains simultaneously. This paper addresses all three gaps through an integrated review, original comparative tables, and a set of governance recommendations spanning AWS and cyber warfare together.

3. Methodology

This paper employs a qualitative, interpretive research design combining doctrinal legal analysis, applied ethics, and a structured comparative review of documented incidents and state positions. Primary sources include United Nations documentation (reports of the Secretary-General, resolutions of the General Assembly First Committee, and reports of the CCW Group of Governmental Experts on LAWS), the Tallinn Manual 2.0, ICRC policy statements, and reports of UN Panels of Experts (notably on Libya). Secondary sources include peer-reviewed articles in international law, military ethics, and security studies journals, together with reports from Human Rights Watch, the Stockholm International Peace Research Institute (SIPRI), and the Campaign to Stop Killer Robots / Stop Killer Robots coalition. Quantitative material — including cybercrime cost estimates and UN voting records — is drawn from named, dated, and publicly available sources and is presented descriptively rather than through original statistical modelling; the paper is not an empirical study in the social-scientific sense but a structured, citation-grounded synthesis intended to meet the evidentiary standards of policy-oriented legal and ethical scholarship. All sources are cited in Vancouver numbered style, with a consolidated reference list at the end of the paper. Where sources disagree — for example, on whether the Kargu-2 episode in Libya constituted the first lethal autonomous engagement — the disagreement is reported rather than resolved by assertion, consistent with the paper's aim of presenting balanced, multi-perspective analysis.

4. Autonomous Weapons Systems: Technology and Typology

There is no single, universally accepted legal definition of an autonomous weapons system. The ICRC defines an AWS functionally, as "any weapon system with autonomy in its critical functions of selecting and attacking targets," a definition that deliberately avoids reference to any particular technology (artificial intelligence, machine learning, or otherwise) so as to remain applicable as the underlying technology evolves [20]. The United States Department of Defense's Directive 3000.09 similarly defines an autonomous weapon system as one that, "once activated, can select and engage targets without further intervention by a human operator," while explicitly distinguishing autonomous systems from semi-autonomous systems in which a human selects the specific target or specific group of targets before the system executes the engagement [3,30].

Scholars commonly distinguish three levels of human involvement in the targeting cycle: "human-in-the-loop" systems, in which a human must authorise each engagement; "human-on-the-loop" systems, in which a machine can act autonomously but a human retains real-time monitoring and override capability; and "human-out-of-the-loop" systems, in which the machine acts with no possibility of

human intervention once activated [21,31]. Existing operational systems span this spectrum. Fully automated defensive systems — such as Israel's Iron Dome, the US Navy's Phalanx Close-In Weapons System, and South Korea's SGR-A1 sentry gun deployed along the Demilitarised Zone — have operated in human-on-the-loop or, in narrow defensive circumstances, human-out-of-the-loop modes for decades against fast-moving munitions such as incoming rockets or artillery shells, a use case widely regarded as less ethically fraught because the "targets" are munitions rather than persons [32]. The more contested category is offensive loitering munitions — sometimes called "kamikaze drones" — exemplified by the Israeli-made Harop, the Turkish Kargu-2, and Russia's Lancet-3, which can loiter over a battlefield, use onboard sensors and, increasingly, machine-learning-based image classification to identify targets, and strike without a continuous human data link [4,33].

Three technical trends are converging to expand the practical reach of autonomy in weapons systems. The first is the maturation of computer-vision-based object and person recognition, which has improved sufficiently for military planners to consider automated target classification feasible in constrained environments, even though independent researchers continue to document high error rates for such systems under battlefield conditions including smoke, camouflage, and unusual poses [16,34]. The second is swarming, in which multiple autonomous platforms coordinate via peer-to-peer communication to overwhelm defences or to distribute targeting decisions across a networked group rather than a single platform — a capability STM has publicly marketed for the Kargu system and that China, Russia, and the United States are all separately developing [21,25]. The third is the integration of large-scale data processing and AI-assisted decision support into command-and-control systems, exemplified by reported Israeli military use of AI-assisted target-generation systems (publicly referred to in investigative reporting as "Lavender" and "The Gospel") during the 2023–2024 Gaza conflict, which illustrates how autonomy can enter the targeting cycle "upstream" of the weapon itself, at the level of target nomination rather than terminal engagement [35].

Timeline of Milestone Events in Cyber Warfare and Autonomous Weapons Development (2007–2025)



Source: compiled by the author from sources cited in Sections 2, 5, and 6 [11,20,23,29,31,37,40].

Figure 1. Timeline of milestone events in cyber warfare and autonomous weapons development, 2007–2025 (compiled by the author from sources cited in Sections 4, 5, and 8).

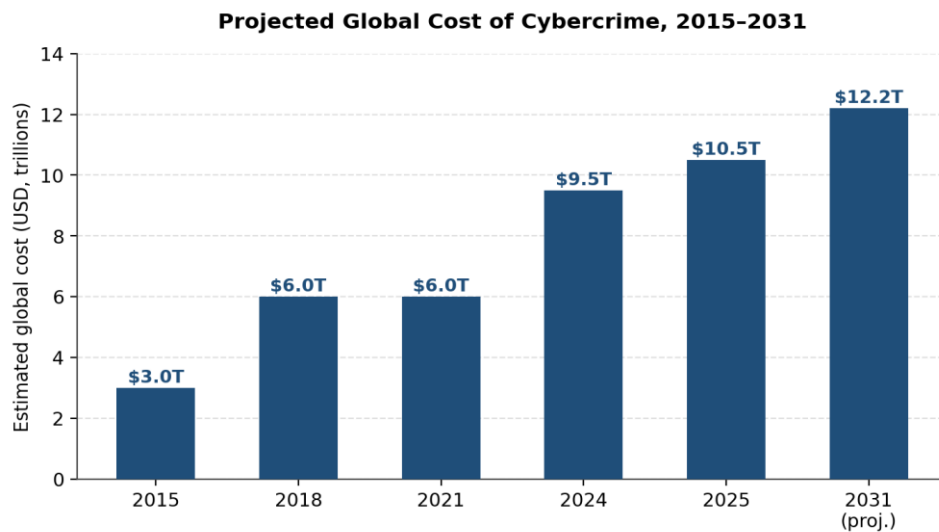
5. Cyber Warfare: Evolution, Characteristics, and Scale

Cyber warfare is best understood not as a single category of action but as a spectrum running from espionage and information operations, through disruptive but non-destructive intrusion (denial-of-service attacks, data theft), to destructive sabotage causing physical or economic damage comparable to kinetic force [26,27]. Three characteristics distinguish cyber operations from conventional or even autonomous kinetic weapons. First, cyber effects are often reversible or contained to data and systems rather than physical structures, which complicates the application of IHL concepts built around physical destruction and casualties [24]. Second, attribution is frequently ambiguous: technical indicators can be spoofed, false flags are common, and definitive public attribution — when it occurs — typically follows a lengthy forensic and intelligence process, as seen in the multi-year attribution efforts surrounding NotPetya and the 2020 SolarWinds supply-chain compromise [28,36]. Third, cyber operations can cascade unpredictably across interconnected systems and borders, meaning an operation targeted at a single organisation (as NotPetya was, initially aimed at Ukrainian tax software) can cause billions of dollars in damage to unrelated third parties in dozens of countries within days [9,10].

Table 1 below summarises six incidents that, taken together, illustrate the evolution of cyber warfare from disruptive but largely reversible action toward operations with substantial destructive, economic, or battlefield significance.

Incident (Year)	Alleged/ Attributed Actor	Primary Effect	Estimated Impact
Estonia DDoS campaign (2007)	Unattributed; widely linked to Russia-based actors	Denial of service against government, banking, and media websites	Weeks of disrupted access to critical online services; first widely cited case of state-level cyber conflict
Stuxnet (2010)	Reportedly a joint US–Israeli operation (never officially confirmed)	Physical sabotage of uranium-enrichment centrifuges at Natanz, Iran	Destroyed roughly 1,000 centrifuges; set back Iran's enrichment programme by an estimated 1–2 years
Sony Pictures hack (2014)	Attributed by the FBI to North Korea	Data theft, destructive wiper malware, coercive intent	Significant financial and reputational damage; rare public state attribution
NotPetya (2017)	Attributed by US, UK, and allied governments to Russia's GRU	Self-propagating wiper malware disguised as ransomware	Approximately US\$10 billion in global damages across shipping, pharmaceuticals, and logistics firms
Kargu-2 loitering munition use, Libya (2020)	Turkish-supplied system used by GNA-aligned forces	Autonomous/near-autonomous engagement of retreating forces	First UN-documented case raising the possibility of an autonomous lethal engagement without operator data-link
SolarWinds supply-chain compromise (2020)	Attributed by US agencies to Russia's SVR	Espionage via compromised software updates	Compromised roughly 18,000 organisations, including multiple US federal agencies

Table 1. Selected milestone incidents illustrating the evolution of cyber warfare and autonomous weapons use, 2007–2020. Compiled by the author from sources cited in Section 5 and Section 8.



Source: Cybersecurity Ventures, Cybersecurity Almanac 2025 [42].

Figure 2. Projected global cost of cybercrime, 2015–2031 (USD, trillions). Source: Cybersecurity Ventures, Cybersecurity Almanac 2025 [11].

The scale illustrated in Figure 2 is instructive for a human-security analysis: cybercrime and cyber-enabled state operations are not analytically separable in practice, since criminal infrastructure (botnets, ransomware-as-a-service platforms, and stolen credential markets) is routinely leased or co-opted by state actors, and state-developed exploits leak into criminal use, as occurred when the US National Security Agency's EternalBlue exploit was leaked and subsequently incorporated into both the WannaCry and NotPetya attacks [9,37]. Cyber warfare therefore cannot be governed as a purely inter-state phenomenon; any credible legal or normative framework must account for the criminal ecosystem that increasingly executes state-directed or state-tolerated operations at arm's length.

6. Ethical Boundaries

Ethical analysis of autonomous weapons and cyber warfare draws on at least three distinct philosophical traditions, each of which yields different conclusions about where the boundary between acceptable and impermissible conduct should lie. Deontological approaches, exemplified by Asaro's dignity argument, hold that certain acts — including the delegation of a lethal decision to a non-human agent — are wrong in themselves, irrespective of consequences [17]. Consequentialist approaches, exemplified by Arkin's defence of "ethical governor" architectures for autonomous systems, evaluate autonomous weapons instrumentally, asking whether their use produces better or worse outcomes (fewer civilian casualties, more consistent IHL compliance) than the human-operated alternative [39]. Virtue-ethics approaches, comparatively underdeveloped in this literature but gaining attention, ask what the increasing reliance on autonomous lethal systems does to the character and moral formation of military institutions and

personnel over time, independent of any single engagement's outcome [31]. This paper does not attempt to adjudicate between these traditions; rather, it uses their coexistence to explain why the ethical debate has proven so durable and resistant to resolution by empirical or technical argument alone — a point developed further below.

6.1 The Accountability (Responsibility) Gap

The most frequently invoked ethical objection to autonomous weapons is the "responsibility gap": when a machine's targeting algorithm produces a wrongful death, it is often unclear whether moral and legal responsibility should attach to the commander who deployed the system, the programmer who designed its target-recognition logic, the manufacturer, or no one at all, because the causal chain between human decision and lethal outcome has been mediated by a system whose precise decision logic may be opaque even to its designers (the "black box" problem of deep-learning systems) [18,38]. This is not merely a theoretical concern: the principle of individual criminal responsibility, foundational to international criminal law since Nuremberg, presumes an identifiable human decision-maker whose mental state (*mens rea*) can be assessed. Autonomous targeting decisions strain this presumption because the "decision" is the emergent output of a statistical model trained on data, not a discrete act of will by an identifiable person at the moment of engagement [17,18].

6.2 Human Dignity and the Right Not to Be Killed by a Machine

A distinct ethical claim, associated most closely with Asaro and echoed in the ICRC's 2021 position, holds that even a hypothetically flawless autonomous weapon — one that discriminates between combatants and civilians with perfect accuracy — would remain morally objectionable because it reduces a human life-and-death decision to a computational output, denying the targeted individual the basic dignity of being killed (or spared) through a deliberative process involving human empathy, context, and moral judgement [17,20]. Critics of this position, including Arkin, counter that human decision-making in combat is itself compromised by fear, fatigue, vengeance, and cognitive bias, and that a sufficiently well-designed autonomous system could in principle exhibit more consistent compliance with IHL's principles of distinction and proportionality than a frightened or angry human soldier [39]. This disagreement — between a deontological objection grounded in the inherent wrongness of machine-mediated killing and a consequentialist defence grounded in potential harm reduction — remains unresolved in the literature and is, this paper argues, unlikely to be resolved by empirical evidence alone, since it turns on contested premises about the moral significance of the decision-making process itself, independent of outcomes.

6.3 Automation Bias and the Erosion of Meaningful Human Control

A growing body of human-factors research documents "automation bias": the well-established tendency of human operators to over-trust automated system outputs, even when those outputs conflict with independently available information, particularly under time pressure [40]. This finding has direct relevance for "human-on-the-loop" systems, which are frequently presented by states as a sufficient safeguard because a human retains formal override authority. If, however, the human overseer is cognitively predisposed to defer to the machine's recommendation — especially in the compressed decision windows characteristic of missile defence or swarm engagements — the formal presence of a human in the approval chain may not translate into meaningful human control in practice. This gap between formal and substantive human control is central to ongoing CCW negotiations over how to define "meaningful human control" in a manner that is operationally verifiable rather than merely declaratory [21,41].

6.4 Ethical Dimensions of Cyber Warfare

Cyber warfare raises related but distinct ethical questions. Just war theory's requirement of discrimination between combatants and civilians is difficult to apply in a domain where the same physical infrastructure (undersea cables, cloud servers, domain name registries) often serves both military and civilian functions simultaneously — the problem of "dual-use" infrastructure [24,27]. Proportionality assessment is similarly complicated by the difficulty of predicting cascading effects in advance: the operators who launched NotPetya could not have fully anticipated its spread to a Cadbury factory in Tasmania, a hospital system in Pennsylvania, or global shipping giant Maersk's terminal operations, yet all suffered substantial harm as unintended consequences of an operation nominally directed at Ukraine [9,10]. Finally, the near-total anonymity available to cyber operators — the attribution problem discussed in Section 5 — creates a distinct ethical hazard: actors may be more willing to authorise operations they believe carry a low probability of being publicly and definitively attributed to them, a moral hazard structurally similar to, but arguably more severe than, the risk-lowering effect that critics attribute to remotely piloted drone warfare [28,42].

6.5 Proliferation, Arms-Race Dynamics, and Strategic Stability

A final ethical dimension, less frequently foregrounded than the dignity and responsibility-gap arguments but increasingly prominent in the strategic-studies literature, concerns the systemic effects of an emerging autonomous-weapons and offensive-cyber arms race on strategic stability. Altmann and Sauer argue that autonomous weapons capable of operating at machine speed compress decision-

making timelines in ways that could destabilise crisis dynamics between nuclear-armed states, since the perceived need to pre-empt an adversary's autonomous strike capability may create incentives for rapid, poorly deliberated escalation that mirror — and could interact dangerously with — long-standing concerns about "launch on warning" postures in nuclear strategy [32]. A structurally similar concern arises in cyber warfare: because offensive cyber capabilities often rely on the same access and tools used for espionage, and because a defender frequently cannot distinguish an intelligence-gathering intrusion from a pre-positioning operation for future sabotage, incentives to interpret ambiguous cyber activity as threatening can themselves become a source of instability, independent of an adversary's actual intent [27,28]. These "meta-level" ethical concerns — that the mere existence and diffusion of a capability may destabilise the broader security environment even when no single use of that capability is individually unlawful — reinforce the case, developed in Sections 9 through 11 of this paper, for governance frameworks that address proliferation and strategic effects, not solely individual instances of use.

7. International Legal Framework

7.1 International Humanitarian Law and the Martens Clause

International humanitarian law, as codified in the four 1949 Geneva Conventions and their 1977 Additional Protocols, applies to autonomous weapons and cyber operations to the same extent it applies to any other means or method of warfare; no state or scholar seriously disputes that IHL governs these domains in principle [22,43]. The core IHL principles most frequently invoked in the AWS debate are distinction (Additional Protocol I, Article 48), which requires parties to a conflict to distinguish at all times between civilians and combatants; proportionality (Article 51(5)(b)), which prohibits attacks expected to cause civilian harm excessive in relation to the concrete and direct military advantage anticipated; and precaution (Article 57), which requires those who plan or decide upon an attack to take feasible precautions to minimise civilian harm [43]. Critics argue that all three principles presuppose a contextual, qualitative judgement that current machine-learning systems cannot reliably replicate, particularly distinction in complex, cluttered, or urban environments where combatant status is not visually self-evident [16,34].

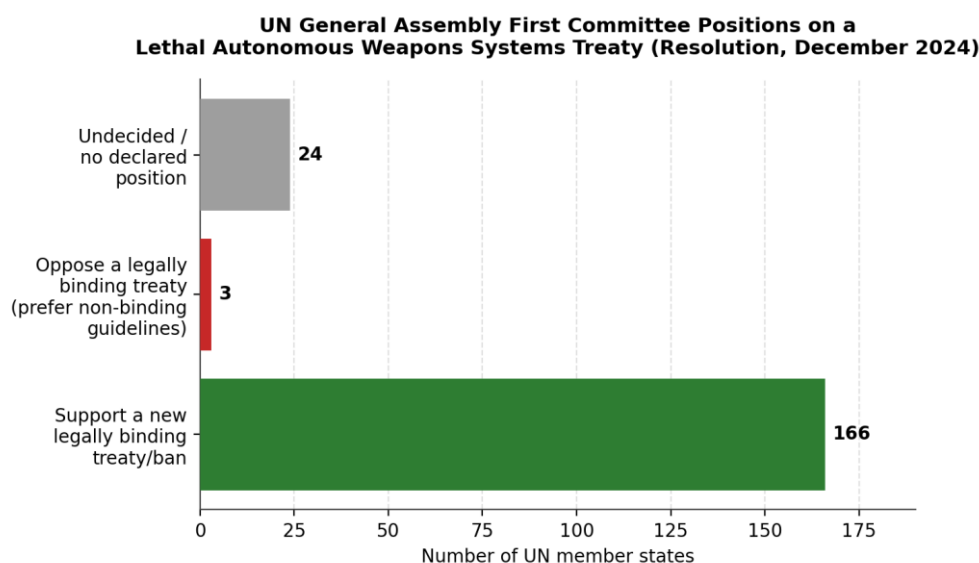
The Martens Clause — first articulated in the 1899 Hague Convention preamble and reaffirmed in Additional Protocol I — provides that in cases not covered by specific treaty provisions, civilians and combatants remain protected by "the principles of humanity and the dictates of public conscience." This clause has become a central rhetorical and legal anchor for prohibitionist advocacy, since campaigners

argue that the delegation of lethal targeting decisions to machines violates the "dictates of public conscience" even in the absence of a specific treaty provision naming autonomous weapons [19,20]. States more sceptical of a binding prohibition, including the United States and Russia, have questioned whether the Martens Clause imposes independently enforceable legal obligations or functions instead as an interpretive principle of last resort [31,41].

7.2 The CCW Process and the Group of Governmental Experts on LAWS

The Convention on Certain Conventional Weapons (CCW) has served as the primary multilateral forum for discussing lethal autonomous weapons systems since 2014, when states parties began holding informal expert meetings; a formal Group of Governmental Experts (GGE) on LAWS was established in 2016 [41,44]. Progress within the GGE has been slow and consensus-bound: the CCW operates by consensus among its roughly 128 states parties, giving any single objecting state an effective veto over binding outcomes [45]. In 2019 the GGE adopted eleven non-binding Guiding Principles, affirming, among other things, that IHL continues to apply fully to potential AWS and that human responsibility for decisions on the use of weapons systems must be retained, since accountability cannot be transferred to a machine [41]. These principles, while useful as a normative baseline, stopped short of the legally binding instrument that a majority of states, the UN Secretary-General, and the ICRC have since called for.

Momentum shifted markedly in 2023–2025. In October 2023, UN Secretary-General António Guterres and ICRC President Mirjana Spoljaric issued a joint call for states to conclude a legally binding instrument on autonomous weapons systems by 2026 [46]. On 2 December 2024, the UN General Assembly adopted its second resolution on LAWS, passing by a vote of 166 in favour to 3 against (Belarus, the Democratic People's Republic of Korea, and the Russian Federation), with 15 abstentions — a striking margin of consensus that nonetheless left the three dissenting votes among states most actively developing autonomous weapons capability [47,48]. In May 2025, the UN General Assembly convened its first plenary meeting dedicated specifically to autonomous weapons systems, with officials from 96 countries participating alongside the ICRC and civil-society organisations, and more than 120 states expressing support for negotiating a treaty [49,50]. Figure 3 summarises the voting pattern of the December 2024 resolution.



Source: UN General Assembly First Committee resolution voting record, 2 December 2024 (166 in favour, 3 against, 15 abstentions, remainder absent/non-voting); compiled from Campaign to Stop Killer Robots and Human Rights Watch reporting [8,9].

Figure 3. UN General Assembly First Committee positions reflected in the vote on the 2024 resolution on lethal autonomous weapons systems (166 in favour, 3 against, 15 abstentions; remaining states absent or non-voting). Compiled by the author from Human Rights Watch and Campaign to Stop Killer Robots reporting [47,48].

7.3 Cyber Operations and the Tallinn Manual 2.0

Unlike AWS, cyber warfare has not been the subject of a dedicated multilateral treaty comparable to the CCW process. Instead, the most authoritative statement of applicable law is the Tallinn Manual 2.0, an unofficial but highly influential restatement produced by an international group of legal experts under NATO Cooperative Cyber Defence Centre of Excellence auspices [22,23]. The Manual affirms that existing international law — including the UN Charter's prohibition on the use of force (Article 2(4)) and the inherent right of self-defence (Article 51), the law of state responsibility, and IHL — applies to cyber operations, but it also identifies substantial areas of unresolved disagreement among experts, including the precise threshold at which a cyber operation constitutes a "use of force" or an "armed attack," and whether non-destructive but highly disruptive operations (for example, against a stock exchange or an electoral system) could ever cross that threshold [24,25]. Complementing the Tallinn process, the UN Group of Governmental Experts (GGE) and, since 2019, the broader Open-Ended Working Group (OEWG) on developments in the field of information and telecommunications have produced consensus reports affirming that international law applies to cyberspace and endorsing eleven voluntary, non-binding norms of responsible state behaviour — including norms against attacking critical infrastructure and against knowingly allowing one's territory to be used for internationally wrongful cyber acts — but, like the CCW Guiding Principles, these remain non-binding and lack an enforcement or verification mechanism [51,52].

7.4 Comparative Assessment of the Two Regimes

Table 2 compares the AWS and cyber-warfare governance regimes across five dimensions relevant to their capacity to constrain harmful state behaviour effectively.

Dimension	Autonomous Weapons Systems (CCW/GGE process)	Cyber Warfare (Tallinn Manual / UN GGE-OEWG process)
Legal status of core instrument	Non-binding Guiding Principles (2019); treaty negotiations proposed but not yet begun	Non-binding expert restatement (Tallinn Manual 2.0) plus non-binding UN norms
Decision rule	Consensus among 128 CCW states parties (single-state veto)	No single negotiating forum; fragmented across UN GGE/OEWG, regional bodies, and expert projects
Attribution/verification mechanism	None; reliant on post-hoc UN Panel of Experts reporting (e.g., Libya)	None binding; attribution is unilateral and politically selective
Major-power engagement	US, Russia, China, Israel, India participate but resist binding prohibition	US, Russia, China, and others accept that law applies in principle but dispute specific thresholds
Trajectory (2023–2025)	Accelerating: UNGA resolutions, Secretary-General 2026 treaty target, first dedicated UNGA meeting (2025)	Comparatively static since Tallinn Manual 2.0 (2017); OEWG renewed but without binding output

Table 2. Comparative assessment of governance regimes for autonomous weapons systems and cyber warfare. Compiled by the author from sources cited in Section 7.

The comparison in Table 2 suggests that AWS governance, while still non-binding, has generated more diplomatic momentum in the 2023–2025 period than cyber-warfare governance, which has remained comparatively static since the publication of the Tallinn Manual 2.0 in 2017. This divergence may reflect the greater visibility and visceral public concern generated by "killer robots" as a category, compared with the more diffuse and often invisible harms of cyber operations, even though the latter's aggregate economic and infrastructural impact, as shown in Figure 2, is almost certainly larger in scale to date.

8. Case Studies: Comparative Analysis

8.1 Libya (2020): The Kargu-2 Episode

In March 2020, during fighting between the UN-recognised Government of National Accord (GNA) and forces aligned with General Khalifa Haftar, the report of the UN Security Council's Panel of Experts on Libya stated that retreating Haftar-affiliated forces and logistics convoys were "hunted down and remotely engaged" by unmanned combat aerial vehicles and lethal autonomous weapons systems, including the STM Kargu-2, which the report described as programmed to attack without requiring a data connection between operator and munition — a "fire, forget and find" capability [4,5]. The manufacturer, STM, has stated that the Kargu-2's autonomous targeting capability has never been demonstrated in combat, and Turkish military sources told journalists that the drones operated autonomously only during transit to the target area, with human operators making the final engagement decision [53]. This disagreement between the UN Panel's characterisation and the manufacturer's and operator's account illustrates a recurring evidentiary problem in AWS governance: verifying autonomous engagement after the fact, from a conflict zone, using forensic and testimonial evidence alone, is extremely difficult, which in turn complicates both legal accountability and policy-making based on documented incidents.

8.2 Nagorno-Karabakh (2020): Loitering Munitions at Scale

The 2020 Nagorno-Karabakh war between Armenia and Azerbaijan is widely regarded by military analysts as the first conflict in which loitering munitions and armed drones played a decisive, rather than supplementary, role in determining battlefield outcomes, with Azerbaijani forces employing Israeli-made Harop loitering munitions and Turkish-made Bayraktar TB2 drones to destroy Armenian armour, artillery, and air-defence systems on a scale that reshaped assessments of the offence-defence balance in conventional warfare [21,54]. While the degree of autonomy exercised by these specific systems in this conflict remains disputed — the Harop, for example, can operate in both autonomous and human-supervised modes — the conflict is frequently cited in the literature as evidence that loitering munitions with autonomous or near-autonomous targeting capability have moved from experimental status to a proven, exportable, and increasingly affordable battlefield capability accessible to a widening range of state and, potentially, non-state actors [54,55].

8.3 Ukraine (2022–present): Mass Adoption and AI-Assisted Targeting

The war in Ukraine since February 2022 has accelerated the diffusion of autonomous and semi-autonomous targeting technology to an unprecedented degree. Both Ukrainian and Russian forces have deployed large numbers of first-person-view (FPV) drones, increasingly incorporating AI-assisted terminal guidance designed to keep the drone locked onto a target even after an electronic-warfare jamming signal severs the operator's control link — a direct technical response to the proliferation of counter-drone jamming on both sides [56]. This development is significant because it demonstrates that autonomy in terminal guidance is emerging not primarily from top-down doctrinal choice but as a bottom-up battlefield adaptation to electronic warfare, a dynamic that existing arms-control frameworks, built around negotiated prohibitions on specific weapon categories, are poorly equipped to anticipate or regulate [21,56]. Reporting on the conflict also indicates that both sides have experimented with computer-vision models trained to recognise armoured vehicles, personnel, and fortifications from drone footage, feeding target coordinates into artillery and strike-drone systems at a tempo that substantially compresses the traditional targeting cycle; while the degree of autonomous engagement (as opposed to AI-assisted human decision support) in these systems remains difficult to verify independently, military analysts broadly agree that the war has functioned as a large-scale, real-world testbed accelerating the maturation of technologies that pre-war arms-control discussions had treated as largely prospective [56]. This has practical implications for the treaty-design recommendations advanced in Section 11: any instrument that focuses solely on a discrete category of "fully autonomous weapons" defined by design intent, rather than on the operationally realised degree of human control at the point of engagement, risks being rendered obsolete by exactly this kind of field-driven, incremental automation.

8.4 NotPetya (2017): Cyber Warfare's Cascading Effects

The NotPetya attack, which US, UK, and allied governments formally attributed to Russia's military intelligence service (the GRU), was initially delivered through a compromised update to a Ukrainian tax-accounting software package, M.E.Doc, and was widely assessed by Western governments as intended to damage Ukrainian government and corporate systems in the context of the ongoing conflict over Crimea and the Donbas [9,57]. Its self-propagating design, however, caused it to spread far beyond Ukraine within hours, disabling systems at shipping conglomerate Maersk, pharmaceutical company Merck, logistics firm FedEx/TNT Express, and a chocolate factory in Tasmania, among many others, with total global damages estimated at approximately US\$10 billion — making it, by most assessments, the most economically destructive single cyber incident publicly documented to date [10,57]. The case

is frequently cited as the clearest evidence that the physical-destruction threshold traditionally used to distinguish "cyber war" from "cyber sabotage" or "cybercrime" is analytically inadequate, since NotPetya caused no direct loss of life but produced economic and operational harm comparable to, or exceeding, that of some conventional military strikes [26,27].

8.5 Stuxnet (2010): The Precedent for Physically Destructive Cyber Sabotage

Although it predates the other cases considered here by roughly a decade, Stuxnet remains the essential reference point for any legal or ethical analysis of destructive cyber operations, because it is the clearest documented instance of a cyber operation causing physical destruction to hardware in a manner functionally equivalent to a kinetic strike. Discovered in 2010 and extensively reconstructed by independent researchers, Stuxnet was a highly sophisticated worm that exploited multiple previously unknown ("zero-day") software vulnerabilities to infiltrate the industrial control systems managing uranium-enrichment centrifuges at Iran's Natanz facility, causing an estimated 1,000 centrifuges to spin at damaging speeds while feeding falsified readings to plant operators to conceal the sabotage [8,9]. Although never officially acknowledged, Stuxnet has been widely reported, including by former US officials speaking anonymously to journalists, to have been a joint operation between the United States and Israel [9]. Legally, Stuxnet raises an unresolved question central to the Tallinn Manual's analysis: because it caused physical damage to Iranian property without loss of life and outside the context of an acknowledged armed conflict between the United States/Israel and Iran, experts remain divided on whether it constituted a "use of force" under Article 2(4) of the UN Charter, an "armed attack" triggering a right of self-defence under Article 51, or a covert act of sabotage governed primarily by the law of state responsibility rather than the law of armed conflict [22,24]. This unresolved classification is emblematic of the broader legal uncertainty documented throughout Section 7.

8.6 Cross-Case Synthesis

Three patterns emerge from the comparative reading of these cases. First, verification and attribution remain the central practical obstacle in both domains: whether the question is "did a machine autonomously select this target?" (Libya) or "who authorised this intrusion?" (NotPetya), after-the-fact forensic reconstruction is difficult, contested, and rarely conclusive enough to support formal legal accountability. Second, both domains exhibit rapid diffusion from a small number of technologically advanced state developers to a widening circle of users, including smaller and middle-power states (Turkey, Azerbaijan) and, plausibly, non-state actors in the future — a trajectory that mirrors the historical proliferation pattern of earlier weapons categories such as man-portable air-defence systems,

but potentially at a faster pace given the lower marginal cost of software-based capability relative to hardware-intensive weapons systems [55,58]. Third, in both domains, the most consequential episodes to date (Libya, NotPetya, Nagorno-Karabakh) occurred in the absence of any binding international legal instrument specifically tailored to the technology involved, underscoring the gap between the pace of technological deployment and the pace of multilateral governance identified throughout this paper.

9. Divergent Stakeholder Perspectives

International governance debates on AWS and cyber warfare are shaped by systematically different interests and risk assessments among four broad stakeholder categories: major military powers with advanced indigenous AWS and cyber capability; middle powers and technology-exporting states; international humanitarian and civil-society organisations; and international organisations and UN bodies. Table 3 presents a structured comparison of their stated positions.

Stakeholder Category	Representative Actors	Core Position on AWS	Core Position on Cyber Norms
Major military powers (AWS/cyber developers)	United States, Russia, China, Israel, United Kingdom	Prefer non-binding guidelines over a treaty; emphasise existing IHL as sufficient; cite strategic necessity and deterrence value	Accept that international law applies but dispute specific thresholds (e.g., "use of force"); resist binding attribution or verification obligations
Middle powers / technology exporters	Turkey, Austria, Brazil, South Korea	Mixed: Turkey exports AWS-capable systems while other middle powers (Austria, Brazil) lead calls for a binding treaty	Generally support UN norms process; limited independent offensive capability shapes more defensive posture
Civil society / humanitarian organisations	ICRC, Human Rights Watch, Campaign to Stop Killer Robots / Stop Killer Robots coalition	Call for a legally binding prohibition on unpredictable AWS and a requirement of meaningful human control over all others	Advocate for protection of civilian infrastructure (hospitals, power grids) from cyber operations under IHL
International organisations	UN Secretary-General's Office, UNGA First Committee, UNIDIR	Actively pushing for treaty conclusion by 2026; convened first dedicated UNGA meeting in 2025	Support OEWG process; limited enforcement authority; rely on member-state consensus

Stakeholder Category	Representative Actors	Core Position on AWS	Core Position on Cyber Norms
Global South states (bloc voting pattern)	Algeria, Costa Rica, Pakistan, Palestine, and 26 others formally calling for a ban since 2013	Among the strongest and most consistent supporters of a binding ban, citing risk of AWS proliferation to conflicts disproportionately affecting the Global South	Generally support binding cyber norms; often lack independent attribution or forensic capacity, increasing reliance on collective mechanisms

Table 3. Comparative stakeholder positions on autonomous weapons systems and cyber-warfare governance. Compiled by the author from sources cited in Sections 7 and 9.

This comparison reveals a structural asymmetry that likely explains the slow pace of binding governance in both domains: the states with the greatest capacity to shape a treaty's technical parameters (the United States, Russia, and China) are also the states with the least strategic incentive to accept binding constraints on capabilities in which they hold a comparative advantage. Conversely, the coalition most vocally demanding a binding instrument — comprising more than thirty Global South states, the ICRC, and civil-society organisations — has moral and numerical weight within UN General Assembly processes (as the 166-vote majority in December 2024 demonstrates) but lacks the capability or leverage to compel binding commitments from the states whose behaviour a treaty would need to constrain most directly [47,48]. This dynamic is analytically similar to earlier arms-control negotiations, including those over landmines and cluster munitions, in which humanitarian-driven treaties (the 1997 Ottawa Treaty and the 2008 Convention on Cluster Munitions) ultimately entered into force without the participation of several of the largest producers and users of the weapons concerned, achieving significant but incomplete normative effect [58,59].

10. Implications for Human Security

Paris cautions that human security, as a concept, risks becoming so expansive — encompassing threats ranging from armed conflict to disease, environmental degradation, and economic insecurity — that it loses analytical precision and becomes difficult to operationalise in policy terms [15]. This paper adopts a narrower operationalisation, focused specifically on the individual-level physical, infrastructural, and informational harms attributable to autonomous weapons and cyber operations, in order to preserve analytical tractability while retaining the paradigm's central insight: that the referent object of security analysis should be the person, household, and community, not the state alone.

The human security framework directs attention away from the question "does this technology threaten state sovereignty or military balance?" and toward the question "does this technology threaten the safety, dignity, and basic needs of individuals and communities?" [14,15]. Applied to autonomous weapons, this reframing highlights concerns that a purely state-security analysis tends to under-weight: the risk of proliferation to non-state armed groups and criminal organisations, given the comparatively low cost and increasing commercial availability of drone and AI components relative to traditional precision-guided munitions [55]; the risk of algorithmic bias in target-recognition systems trained on non-representative data, potentially producing discriminatory targeting patterns that disproportionately affect specific ethnic, religious, or demographic groups [34,35]; and the chilling effect of persistent autonomous surveillance-strike capability on civilian freedom of movement and assembly in contested areas, independent of whether any specific strike occurs [29].

Applied to cyber warfare, the human security lens foregrounds the routine targeting of civilian critical infrastructure — hospitals, water-treatment systems, and power grids — which has occurred both incidentally (as an unintended cascading effect, as in NotPetya's disruption of hospital systems in the United States) and, in several documented cases during the Russia-Ukraine conflict, as an apparent direct target of military and intelligence-linked cyber operations against Ukrainian energy infrastructure [9,60]. It also foregrounds the disproportionate vulnerability of low- and middle-income countries, which frequently lack the cybersecurity budgets, technical expertise, and forensic capacity of wealthier states, making them both more attractive targets for opportunistic criminal exploitation and less able to attribute or respond to state-sponsored intrusions [11,61]. The 2017 WannaCry ransomware attack, which used the leaked NSA EternalBlue exploit and disrupted roughly a third of England's National Health Service trusts, cancelling an estimated 19,000 medical appointments, is frequently cited in the human-security literature as the paradigmatic case of a cyber operation whose most severe harm fell not on military or government targets but on ordinary patients and healthcare workers [37,62].

Taken together, the case for a human-security reframing of AWS and cyber-warfare governance rests on the observation that the harms most likely to materialise at scale — proliferation to poorly regulated actors, algorithmic bias, disruption of civilian critical infrastructure, and disproportionate impact on states and populations with the least capacity to defend themselves — are precisely the harms that a state-centric, great-power-focused arms-control paradigm is least well designed to detect, measure, or prevent. This does not imply that state-security considerations are irrelevant; strategic stability among major powers remains a necessary, if insufficient, condition for effective governance, since a treaty that collapses great-power cooperation entirely would likely produce worse humanitarian outcomes than incremental, imperfect regulation. Rather, the human security frame is best understood as a

complementary lens that corrects for the systematic blind spots of traditional security analysis, ensuring that the interests of civilians, patients, and communities in the Global South — who are simultaneously the most exposed to the proliferation risks documented in this section and the least represented in the great-power negotiations that will determine the shape of any future treaty — are not treated as a secondary consideration in the design of governance instruments for autonomous weapons and cyber warfare.

11. Discussion and Policy Recommendations

Building on the analysis above, this paper advances four interlinked policy recommendations.

First, negotiations toward a legally binding instrument on autonomous weapons systems, as called for by the UN Secretary-General and the ICRC for conclusion by 2026, should adopt a two-tier structure: an outright prohibition on autonomous weapons that cannot be operated under meaningful, verifiable human control or whose effects cannot be reliably predicted by the deploying party, combined with a binding requirement of meaningful human control — defined in operationally testable terms rather than aspirational language — for all other systems with autonomous targeting functions [20,41,46]. This tiered approach reflects the emerging convergence between the ICRC's position and the working papers submitted by states such as Austria, while remaining realistic about the unwillingness of major military powers to accept a blanket ban on defensive systems such as Iron Dome or Phalanx.

Second, cyber-warfare governance should move beyond the current reliance on voluntary, non-binding norms toward a binding framework establishing minimum standards for the protection of designated categories of civilian critical infrastructure — hospitals, water systems, and power grids — analogous to the special protections IHL already affords to medical facilities and personnel in kinetic conflict, alongside a multilateral, technically resourced attribution mechanism modelled on (but more resourced than) the UN Panel of Experts model used for Libya, to reduce the current dependence on unilateral, and therefore politically contestable, state attribution [24,28,51].

Third, both regimes should incorporate mandatory pre-deployment testing, certification, and incident-reporting requirements analogous to those used in civil aviation and pharmaceutical safety regulation, shifting the evidentiary burden away from post-hoc UN Panel of Experts investigations (as in Libya) and toward ex ante verification that a system's autonomous functions meet agreed technical and legal standards before fielding [39,41].

Fourth, given the demonstrated asymmetry between the humanitarian coalition's normative influence and its lack of leverage over major developers (Section 9), civil society and middle-power diplomacy should draw explicitly on the precedent of the Ottawa and Oslo processes for landmines and cluster munitions: pursuing a treaty open for signature even without universal major-power participation, on the premise that even a partial, non-universal treaty can shift global norms, stigmatise non-compliant behaviour, and constrain the commercial export of the most dangerous system categories over time [58,59].

These recommendations are not without limitations. A tiered prohibition requires a technically robust and internationally agreed definition of "meaningful human control," which the CCW process has debated for over a decade without resolution; a binding cyber-infrastructure-protection regime requires attribution capacity that many states currently lack; and any treaty lacking major-power participation will, as with landmines and cluster munitions, leave the most capable and most frequently involved actors outside its direct legal reach. These limitations do not, in this paper's assessment, outweigh the risks of continued inaction, but they do mean that governance progress in both domains is likely to remain incremental, contested, and dependent on sustained coalition-building between international organisations, civil society, and the growing number of middle-power and Global South states willing to lead by example.

A fifth, cross-cutting recommendation concerns the treatment of dual-use and criminal infrastructure identified in Section 5. Because state-sponsored cyber capability is increasingly intertwined with, and sometimes indistinguishable from, criminal ransomware and botnet infrastructure, effective governance cannot be achieved through inter-state treaty-making alone; it also requires sustained investment in the cybersecurity capacity of low- and middle-income states, coordinated law-enforcement action against criminal infrastructure providers regardless of the jurisdiction in which they operate, and mandatory vulnerability-disclosure norms for the private technology companies whose products form the substrate of both defensive and offensive cyber operations [11,61]. Absent this capacity-building dimension, a binding norms framework negotiated primarily among technologically advanced states risks reproducing, in cyberspace, the same disparity in practical protection that already characterises the physical-security gap between well-resourced and under-resourced states.

11.1 Limitations of the Present Study

Several limitations qualify the conclusions offered above. First, this paper relies substantially on open-source, publicly reported information regarding state cyber operations and autonomous-weapons

deployments; classified assessments held by national intelligence and defence agencies may present a materially different picture of both the scale of activity and the technical sophistication of autonomy actually exercised in specific incidents, including the Libya and Nagorno-Karabakh cases discussed in Section 8, where manufacturer and operator accounts diverge from UN Panel of Experts characterisations. Second, the pace of technological change in both domains is rapid enough that specific technical claims — for example, regarding the current state of AI-assisted target recognition or terminal-guidance resilience to jamming — risk becoming dated within a short period following publication; the analysis should therefore be read as characterising trends and structural dynamics rather than a fixed technical snapshot. Third, the paper's comparative tables (Tables 1 through 3) necessarily compress complex, contested, and in some cases actively disputed factual records into summary form for analytical clarity; readers seeking granular detail on any individual incident are directed to the primary sources cited in the reference list. Finally, as an interdisciplinary synthesis spanning law, ethics, and security studies, the paper does not claim the methodological rigour of a dedicated empirical study in any one of these fields; its contribution lies instead in the integration of perspectives that are more commonly treated separately, and in the original comparative tables and figures developed to support that integration.

12. Conclusion

Autonomous weapons systems and cyber warfare represent two converging expressions of a single underlying transformation in the conduct of armed conflict: the delegation of consequential, often irreversible decisions to machines and code operating at speeds and scales that increasingly exceed meaningful real-time human oversight. This paper has shown that the ethical objections to this transformation — the accountability gap, the dignity argument, and the empirical reality of automation bias — are substantial and, to date, unresolved by either technological improvement or diplomatic consensus. It has shown that existing international law, while formally applicable to both domains, was not designed with either technology in mind, leaving significant interpretive gaps around distinction, proportionality, attribution, and the threshold for an "armed attack" in cyberspace. Original comparative analysis of state positions, milestone incidents, and stakeholder perspectives has demonstrated that governance momentum, while real and accelerating — particularly for AWS since 2023 — remains structurally constrained by the same asymmetry that has shaped arms control for decades: the states best positioned to develop and deploy these technologies are least incentivised to accept binding limits on them. Reframing the debate around human security rather than state security alone does not resolve this asymmetry, but it does clarify what is genuinely at stake: not an abstract contest of strategic

advantage between great powers, but the safety, dignity, and basic security of the civilians, patients, and communities who bear the disproportionate cost when autonomous weapons and cyber operations go wrong. As the international community approaches the UN Secretary-General's proposed 2026 target for concluding a binding instrument on autonomous weapons, and as cyber-warfare's economic and infrastructural toll continues to climb toward the trillions of dollars documented in this paper, the central policy challenge is no longer whether governance is necessary, but whether it can be built quickly enough, and inclusively enough, to keep pace with the technology it seeks to constrain.

References

1. United Nations Institute for Disarmament Research (UNIDIR). The Weaponization of Increasingly Autonomous Technologies: Considering how Meaningful Human Control Might Move the Discussion Forward. Geneva: UNIDIR; 2014.
2. Kaldor M. New and Old Wars: Organized Violence in a Global Era. 3rd ed. Stanford: Stanford University Press; 2012.
3. US Department of Defense. Directive 3000.09: Autonomy in Weapon Systems. Washington, DC: DoD; 2012 (updated 2023).
4. National Public Radio. A U.N. report suggests Libya saw the first battlefield killing by an autonomous drone. Washington, DC: NPR; 2021 Jun 1.
5. United Nations Security Council. Letter dated 8 March 2021 from the Panel of Experts on Libya established pursuant to resolution 1973 (2011) addressed to the President of the Security Council. S/2021/229. New York: UN; 2021.
6. Kallenborn Z. Was a Flying Killer Robot Used in Libya? Quite Possibly. Bulletin of the Atomic Scientists. 2021 Apr 20.
7. Scharre P. Army of None: Autonomous Weapons and the Future of War. New York: W.W. Norton & Company; 2018.
8. Herzog S. Revisiting the Estonian Cyber Attacks: Digital Threats and Multinational Responses. Journal of Strategic Security. 2011;4(2):49-60.
9. Zetter K. Countdown to Zero Day: Stuxnet and the Launch of the World's First Digital Weapon. New York: Crown Publishers; 2014.
10. Greenberg A. The Untold Story of NotPetya, the Most Devastating Cyberattack in History. Wired. 2018 Aug 22.
11. Greenberg A. Sandworm: A New Era of Cyberwar and the Hunt for the Kremlin's Most Dangerous Hackers. New York: Doubleday; 2019.
12. Cybersecurity Ventures. Cybersecurity Almanac 2025: 100 Facts, Figures, Predictions and Statistics. Sausalito, CA: Cybersecurity Ventures / Evolution Equity Partners; 2025 Dec 11.

13. Brundage M, Avin S, Clark J, et al. The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Future of Humanity Institute et al.; 2018.
14. Boulanin V, Verbruggen M. Article 36 Reviews: Dealing with the Challenges Posed by Emerging Technologies. Stockholm: SIPRI; 2017.
15. United Nations Development Programme. Human Development Report 1994: New Dimensions of Human Security. New York: Oxford University Press; 1994.
16. Paris R. Human Security: Paradigm Shift or Hot Air? International Security. 2001;26(2):87-102.
17. Sharkey N. Saying 'No!' to Lethal Autonomous Targeting. Journal of Military Ethics. 2010;9(4):369-383.
18. Asaro P. On Banning Autonomous Weapon Systems: Human Rights, Automation, and the Dehumanization of Lethal Decision-Making. International Review of the Red Cross. 2012;94(886):687-709.
19. Sparrow R. Killer Robots. Journal of Applied Philosophy. 2007;24(1):62-77.
20. Human Rights Watch; International Human Rights Clinic, Harvard Law School. Losing Humanity: The Case Against Killer Robots. New York: Human Rights Watch; 2012.
21. International Committee of the Red Cross. ICRC Position on Autonomous Weapon Systems. Geneva: ICRC; 2021 May.
22. Scharre P, Horowitz MC. An Introduction to Autonomy in Weapon Systems. Washington, DC: Center for a New American Security; 2015.
23. Schmitt MN, ed. Tallinn Manual 2.0 on the International Law Applicable to Cyber Operations. 2nd ed. Cambridge: Cambridge University Press; 2017.
24. NATO Cooperative Cyber Defence Centre of Excellence. International Law Applies to Cyber Operations, Tallinn Manual 2.0 Reaffirms. Tallinn: CCDCOE; 2017.
25. Schmitt MN, ed. Tallinn Manual on the International Law Applicable to Cyber Warfare. Cambridge: Cambridge University Press; 2013.
26. Vihul L. The Tallinn Manual 2.0 and the Evolving Law of Cyber Operations. In: Cambridge International Law Journal blog. Cambridge; 2017.
27. Rid T. Cyber War Will Not Take Place. Journal of Strategic Studies. 2012;35(1):5-32.
28. Lin H, Zegart A, eds. Bytes, Bombs, and Spies: The Strategic Dimensions of Offensive Cyber Operations. Washington, DC: Brookings Institution Press; 2019.
29. Rid T, Buchanan B. Attributing Cyber Attacks. Journal of Strategic Studies. 2015;38(1-2):4-37.
30. Suchman L, Weber J. Human-Machine Autonomies. In: Bhuta N, Beck S, Geis R, et al., eds. Autonomous Weapons Systems: Law, Ethics, Policy. Cambridge: Cambridge University Press; 2016. p.75-102.
31. Horowitz MC. The Ethics and Morality of Robotic Warfare: Assessing the Debate over Autonomous Weapons. Daedalus. 2016;145(4):25-36.
32. Altmann J, Sauer F. Autonomous Weapon Systems and Strategic Stability. Survival. 2017;59(5):117-142.

33. Davison N. A Legal Perspective: Autonomous Weapon Systems under International Humanitarian Law. UNODA Occasional Papers. 2017;30:5-18.
34. Amoroso D, Tamburrini G. Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues. Current Robotics Reports. 2020;1:187-194.
35. Abraham Y, Davidson H. 'The Gospel': how Israel uses AI to select bombing targets in Gaza. The Guardian. 2023 Dec 1.
36. US Cybersecurity and Infrastructure Security Agency (CISA). Advisory AA20-352A: Advanced Persistent Threat Compromise of Government Agencies, Critical Infrastructure, and Private Sector Organizations. Washington, DC: CISA; 2020 Dec 17.
37. National Audit Office (UK). Investigation: WannaCry Cyber Attack and the NHS. London: NAO; 2017 Oct 27.
38. Roff HM. The Strategic Robot Problem: Lethal Autonomous Weapons in War. Journal of Military Ethics. 2014;13(3):211-227.
39. Arkin RC. Governing Lethal Behavior in Autonomous Robots. Boca Raton: CRC Press; 2009.
40. Cummings ML. Automation Bias in Intelligent Time Critical Decision Support Systems. AIAA Intelligent Systems Technical Conference. 2004.
41. Convention on Certain Conventional Weapons Group of Governmental Experts on Lethal Autonomous Weapons Systems. Guiding Principles Affirmed by the Group of Governmental Experts. Geneva: UN; 2019.
42. Kaag J, Kreps S. Drone Warfare. Cambridge: Polity Press; 2014.
43. International Committee of the Red Cross. Commentary on the Additional Protocols of 8 June 1977 to the Geneva Conventions of 12 August 1949. Geneva: ICRC; 1987.
44. United Nations Office for Disarmament Affairs. Background on Lethal Autonomous Weapons Systems in the CCW. New York: UNODA; 2023.
45. Geiss R, Lahmann H. Autonomous Weapons Systems: A Paradigm Shift for the Law of Armed Conflict? In: Ohlin JD, ed. Research Handbook on Remote Warfare. Cheltenham: Edward Elgar; 2017. p.371-404.
46. United Nations Secretary-General; International Committee of the Red Cross. Joint Call for Action on Autonomous Weapons Systems. New York/Geneva: UN/ICRC; 2023 Oct.
47. Human Rights Watch. UN: Start Talks on Treaty to Ban 'Killer Robots'. New York: HRW; 2025 May 21.
48. Stop Killer Robots. 161 States Vote Against the Machine at the UN General Assembly. Geneva: Stop Killer Robots; 2024 Dec.
49. Human Rights Watch. UN General Assembly First Committee Approves Second Resolution on 'Killer Robots'. New York: HRW; 2024 Nov.
50. Stop Killer Robots. The Story So Far: Second GGE Meeting of 2025 on Autonomous Weapons. Geneva: Stop Killer Robots; 2025.
51. United Nations Group of Governmental Experts on Developments in the Field of Information and Telecommunications in the Context of International Security. Final Report. New York: UN; 2021.

52. United Nations Open-Ended Working Group on Security of and in the Use of Information and Communications Technologies. Final Substantive Report. New York: UN; 2021.
53. Al-Monitor. Turkish Drone Sets International Buzz Over Killer Robots. Washington, DC: Al-Monitor; 2021 Jun.
54. Shaikh S, Rumbaugh W. The Air and Missile War in Nagorno-Karabakh: Lessons for the Future of Strike and Defense. Washington, DC: Center for Strategic and International Studies; 2020.
55. Fuhrmann M, Horowitz MC. Droning On: Explaining the Proliferation of Unmanned Aerial Vehicles. International Organization. 2017;71(2):397-418.
56. Institute for the Study of War. Ukraine Conflict Updates: Drone Warfare and Electronic Countermeasures. Washington, DC: ISW; 2024-2025.
57. US Department of Justice. Six Russian GRU Officers Charged in Connection with Worldwide Deployment of Destructive Malware and Other Disruptive Actions in Cyberspace. Washington, DC: DOJ; 2020 Oct 19.
58. International Campaign to Ban Landmines. Landmine Monitor 2023. Geneva: ICBL-CMC; 2023.
59. Cluster Munition Coalition. Cluster Munition Monitor 2023. Geneva: CMC; 2023.
60. Council on Foreign Relations. Cyberattack on Ukraine's Power Grid. Washington, DC: CFR Cyber Operations Tracker; 2016 (updated 2022).
61. World Bank. Digital Development Partnership: Cybersecurity Capacity in Developing Countries. Washington, DC: World Bank; 2022.
62. National Health Service Digital. Lessons Learned Review of the WannaCry Ransomware Cyber Attack. London: NHS Digital; 2018.