Ethics and International Law for AI Warfare

MixCache.com

Table of Contents

Introduction

Artificial intelligence is reshaping the character of warfare without altering its tragic nature. From autonomous navigation and target recognition to decision-support at machine speed, these technologies promise new forms of capability, coordination, and risk. They also raise urgent questions that cannot be left to code alone: What moral

principles should guide the design and use of AI-enabled systems? Which legal norms constrain their deployment, and how do those norms apply when systems learn, adapt, or act with limited human oversight? This book responds to those questions by bridging ethical reasoning with international law, offering a clear, normative guide for lawyers, ethicists, and military leaders who must translate values into policy, practice, and compliance.

Our starting point is familiar yet freshly tested terrain: just war principles and the law of armed conflict. Distinction, proportionality, necessity, and precautions in attack remain the bedrock of lawful and legitimate force. But AI systems complicate their application. When algorithms detect targets, when predictive models estimate collateral effects, or when autonomy affects the pace of engagements, the meaning of "reasonable commander judgment" and "feasible precautions" demands concrete, operational interpretation. The Martens Clause and customary international law provide a moral backstop where treaty law is silent, reminding us that humanity and public conscience continue to matter even as technology evolves. The challenge is to translate these enduring commitments into requirements that guide architectures, training, and accountability.

Throughout the book, we treat AI in warfare as a socio-technical system, not a standalone artifact. Capabilities emerge from interactions among data, models, sensors, communications, human operators, and command authority. Degrees of autonomy vary across the kill chain—from detection and tracking to identification, engagement, and assessment—making "meaningful human control" a design and governance question rather than a slogan. We therefore connect ethical considerations to lifecycle obligations: requirements definition, dataset governance, assurance and testing, deployment constraints, monitoring, and post-incident investigation. This lifecycle perspective grounds legal compliance in practical mechanisms, not merely in doctrinal assertions.

Three cross-cutting issues receive particular attention. First, proportionality under uncertainty: how to quantify and bound expected incidental harm when model performance is probabilistic and context-sensitive, and how to preserve commander discretion without outsourcing moral judgment to a threshold in software. Second, attribution and responsibility: when multiple actors—developers, commanders, coalition partners, and even adversaries exploiting vulnerabilities—shape an outcome, who bears legal responsibility and on what theory of fault or state responsibility. Third, command responsibility in the age of learning systems: what counts as "effective control," what meets the standards of foreseeability and due diligence, and how training, supervision, and review processes must adapt when systems update or degrade in the field.

The book also addresses the evidentiary backbone of accountability. Legal and ethical evaluation requires reasons that can be examined after the fact. That means

auditability by design, disciplined logging, model version control, test coverage linked to operational scenarios, and clear chains of custody for digital evidence. Explainability here is not an abstract research aspiration; it is central to weapons reviews, proportionality assessments, rules of engagement, and fair adjudication. Where perfect interpretability is unavailable, we emphasize assurance arguments, performance envelopes, and governance controls that make residual risk explicit and acceptable in light of legal obligations.

Strategic considerations frame the operational details. AI can accelerate tactical timelines, compress decision windows, and create new couplings across domains—from cyber effects that manipulate sensors to swarms that saturate defenses. These dynamics affect escalation management and crisis stability, particularly where misidentification or adversarial deception can produce rapid, correlated errors. We examine pathways to mitigate these risks through doctrine, signaling, technical safeguards, and arms control initiatives, while remaining attentive to the realities of great-power rivalry and coalition operations.

This is a practical book for practitioners. Each chapter connects normative principles to decision points faced by program managers, staff judge advocates, commanders, and oversight bodies. Our aim is neither to celebrate nor to condemn AI in the abstract, but to equip responsible actors to meet existing legal standards and ethical expectations. Where the law leaves gaps, we offer defensible proposals for policy, training, and international cooperation that reduce civilian harm, clarify accountability, and sustain legitimacy.

Finally, we recognize that consensus will not form overnight. Reasonable people disagree about the proper balance between military necessity and humanitarian protection, about the sufficiency of current law, and about the prudence of deploying certain autonomous capabilities at all. Yet the urgency of real-world decisions requires a disciplined framework now. By integrating moral principles with legal norms—and by tying both to concrete governance practices—we seek to make AI-enabled warfare, if it occurs, more accountable, more discriminating, and more humane.

CHAPTER ONE: The Landscape of AI in Armed Conflict

The sound of warfare has always been a symphony of human action, from the thud of boots on unforgiving earth to the roar of cannons and the shouts of command. But the composition is changing, with the silent hum of algorithms and the whir of autonomous systems now joining the chorus. Artificial intelligence, once a realm of

science fiction, has firmly planted itself on the battlefield, transforming the character of conflict in ways that demand our immediate and thoughtful attention. This isn't just about faster computers or smarter bombs; it's about a fundamental shift in how decisions are made, how force is applied, and ultimately, who or what bears responsibility.

Nations across the globe are engaged in a brisk competition, some might even say a race, to integrate AI into their military operations. Ukraine and Russia, for instance, are at the forefront, actively developing and deploying autonomous systems for battlefield advantage. The rapid progress of AI is reshaping modern warfare, influencing everything from strategic planning to battlefield operations and even the ethical considerations surrounding the use of military force. As these nations embed AI into combat, a critical question arises: how much reliance is too much, and at what cost?

The implications are far from theoretical. Austrian Foreign Minister Alexander Schallenberg aptly described the current moment as "the Oppenheimer moment of our generation," drawing a parallel to the transformative impact of nuclear weapons in the 20th century. AI-enabled weapons are similarly reshaping battlefields, with the conflict in Ukraine serving as a stark contemporary example. Schallenberg warned that AI-driven warfare could trigger an uncontrollable arms race, where autonomous drones and algorithm-driven targeting systems could make mass killing a mechanized, almost effortless process.

Indeed, the Ukraine conflict has become a crucible for AI in warfare. Both sides are deeply invested in an AI-driven drone race, leveraging autonomous technologies. Drones now account for a significant portion—roughly 70-80%—of battlefield casualties in the Russia-Ukraine war. Ukraine, facing numerical superiority from Russia, embraced drones early on, prompting Moscow to follow suit. General Valerii Zaluzhnyi, Ukraine's former commander-in-chief, observed that many of Ukraine's drones utilize commercial components and open-source software, facilitating a low-cost attrition warfare strategy.

This intense competition between electronic warfare capabilities and drone operators has spurred rapid innovation. Both sides have developed countermeasures, such as using fiber-optic cables to bypass jamming, and are constantly working on new adaptations. The next evolution in drone warfare is already emerging: AI-powered targeting systems that enable drones to identify and strike targets with minimal human intervention, even in environments heavily affected by jamming. This ongoing struggle for drone supremacy is pushing both Ukraine and Russia to pursue technological breakthroughs relentlessly, potentially transforming warfare into a battle of algorithms.

The deployment of autonomous drones has revolutionized surveillance, reconnaissance, and combat strategies. Generative AI further enhances these

capabilities by enabling real-time decision-making and coordination without direct human input. These AI systems empower drones to adapt to dynamic environments, identify targets, and coordinate with other units to execute missions effectively. In swarm operations, for example, generative AI allows multiple drones to collaborate, sharing information and adjusting tactics in response to threats.

Beyond the immediate theater of conflict, the U.S. military has also been actively integrating AI for many years, even before its widespread adoption in civilian life. The Department of Defense has allocated at least $75 billion to AI-driven programs since 2016, a figure that likely underestimates the true investment due to classified projects and those with unclear AI integration. This funding supports a wide array of applications, including surveillance, targeting, and the development of autonomous weapons capable of selecting targets and taking lethal action with varying levels of human involvement.

One prominent example is the Maven Smart System, a flagship Pentagon AI initiative that has evolved over a decade of collaboration between the Defense Department and the tech industry. Maven aims to improve intelligence analysis, surveillance, and targeting by sifting through vast quantities of information from satellites, data brokers, military drones and sensors, and social media to identify persons and objects of interest. The system also incorporates advanced AI models to accelerate target analysis, generate intelligence, and simulate battlefield scenarios.

Other key players in the defense industry are also making significant strides. Companies like Palantir and Anduril have seen substantial growth in their defense revenue, driven by AI-powered solutions. Palantir, the lead contractor for the Maven Smart System, has deployed its technology in various conflict zones, including Iran, Iraq, Syria, Ukraine, and Yemen. Anduril specializes in autonomous systems like drones and surveillance towers, and develops technology to counter adversary autonomous weapons. Their drones, powered by proprietary AI, can navigate hostile environments, communicate with each other, and potentially engage targets with minimal human involvement.

However, the rapid embrace of AI by the military is not without its challenges and concerns. One significant issue is the risk of over-reliance on technology, which could displace crucial human expertise and judgment in life-or-death decisions, potentially endangering both troops and civilians. As anyone who has interacted with AI chatbots knows, these systems can make mistakes, some obvious, others more subtle and difficult to detect.

In a military context, AI's propensity for inaccuracy can have deadly consequences. For instance, in 2024, Maven's algorithms reportedly identified a tank correctly only about 60% of the time in good weather, with accuracy plummeting to 30% in snowy conditions. Furthermore, advanced AI models can generate persuasive yet false or

misleading analyses, increasing the likelihood that commanders and analysts might accept incorrect recommendations, especially in the high-stress environment of combat. This means that even with humans making final decisions, reliance on AI for target selection or justification can lead to tragic errors.

The military's reliance on commercial technology also presents potential pitfalls. The opacity of proprietary targeting algorithms, for example, makes it challenging for the military to inspect them for inherent biases that might lead to the misidentification of civilians as military objectives. In 2025, the army reportedly flagged a battlefield communications system designed by Palantir and Anduril as a "black box," raising concerns about unauthorized access to its applications and data.

Despite these concerns, the benefits of AI in military applications are also significant and continue to drive its adoption. AI can enhance operational capabilities by improving situational awareness, accelerating decision-making, and enabling more accurate targeting, thereby contributing to the overall effectiveness of military operations. Generative AI, for example, can contribute to strategic decision-making by rapidly sorting through vast amounts of data, identifying connections, patterns, and potential implications that would take humans much longer to uncover.

This information can be presented to human decision-makers not only as reports but also in a conversational format, fostering human-AI collaboration. AI can also create simulations to test various scenarios, leading to more informed decisions. These capabilities are crucial in an increasingly complex and multi-domain security environment, where traditional decision-making approaches often struggle to keep pace with the speed and scale of contemporary threats.

Beyond decision support, AI is being integrated into a multitude of specific military functions. These include intelligent decision-support systems and aided target recognition, which can reduce the mental load on human operators and enable faster responses. This approach offers advantages such as rapid response times, the ability to operate in high-risk environments, and a reduced risk to human personnel.

Cybersecurity is another critical area where AI plays a dual role. While AI can be used by malicious actors to create malware or facilitate social engineering attacks, it also offers powerful tools for defense. Generative AI, with its ability to analyze large datasets and identify patterns, can detect potential threats and use predictive analytics to anticipate future attacks, thereby bolstering cyber defense for critical military applications. The military must continually adapt its training and mitigation plans to counter the evolving threats posed by AI in the hands of adversaries.

Logistics and resupply, often overlooked but foundational to successful military operations, are also being transformed by AI. Generative AI optimizes supply chain routes, forecasts demand, and simulates resupply scenarios, particularly in austere or

contested environments. These models can process real-time data on terrain, weather, and enemy movements to generate efficient resupply plans that minimize risk and maximize speed. This has led to advancements in autonomous resupply systems using unmanned vehicles or drones that can navigate complex environments.

The future battlefield, therefore, will likely be characterized by a growing integration of AI across all aspects of warfare. Military dominance may increasingly be defined by the performance of algorithms rather than solely by the size of an army. The combination of enhanced sensors, automation, and AI with advanced technologies like hypersonics will produce weapons that are more accurate, better connected, faster, and more destructive. These advanced capabilities, while initially concentrated in the most advanced militaries, are expected to proliferate over time, making more assets vulnerable and heightening the risk of escalation.

This evolving landscape underscores the urgent need for a clear normative guide that bridges ethical reasoning with international law. As AI systems become more sophisticated and autonomous, the traditional frameworks for understanding conflict, responsibility, and accountability are being profoundly tested. The challenge lies in translating enduring moral and legal commitments into practical requirements that can guide the design, deployment, and oversight of AI in warfare, ensuring that humanity and public conscience remain central even as technology advances at an unprecedented pace.
