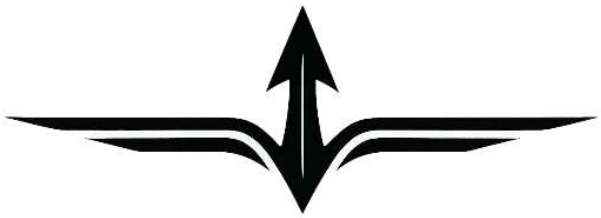
Chapter 16



Responsibility in Machine-Age Warfare

Lukas Filler

"Responsibility is a unique concept…You may delegate it, but it is still with you. You may disclaim it, but you cannot divest yourself of it."[2]

— Admiral Hyman G. Rickover
Testimony on Radiation Safety and Regulation (1961)

Introduction: Morality Under Machine Pressure

Contemporary debates about military artificial intelligence (AI) are often pulled toward familiar but unhelpful extremes.[3] At one end are speculative fears of runaway machines and apocalyptic scenarios in which super-intelligent systems turn against their creators.[4] At the other end lies a technocratic confidence that war can be reduced to an optimization problem, solved through faster processing and better

doi.org/10.71236/g14xbakm

data.[5] Both perspectives distract from the more immediate and consequential danger facing modern militaries. The most pressing ethical risk is not that machines will acquire independent moral standing, but that human commanders will allow these systems to become a pretext for surrendering their own responsibility for judgment.[6]

This drift toward automation is not occurring in isolation. It is driven by acute strategic pressure, particularly the perception of intensifying competition with technologically capable adversaries.[7] China's pursuit of what it calls "intelligentized warfare"—the integration of AI, data fusion, and cognitive operations to compress decision cycles beyond human reaction times—has sharpened fears that hesitation itself may become a strategic liability.[8] For U.S. leaders, the fear of falling behind an adversary's decision tempo increasingly frames speed not as a tradeoff against control, but as a necessity for survival.[9] There is strong potential for this to create a pattern of accelerated deployment in which systems are fielded before their limits are fully understood, and where the urgency to keep pace outstrips the rigor of testing, validation, and ethical reflection.[10]

Yet much of the ethical anxiety surrounding military AI is misplaced. While the character of war—its tempo, reach, and technical sophistication—is clearly changing,[11] the nature of war as a violent human endeavor has not.[12] The dilemmas raised by AI-enabled warfare are not wholly new. They are an intensification of a long-standing trajectory in which technology has progressively distanced those who wield force from those who bear its consequences.[13] From long-range artillery to strategic bombing and standoff precision weapons, modern militaries have repeatedly sought ways to extend lethality while reducing physical and psychological proximity to

harm. Artificial intelligence accelerates this trend, but it does not fundamentally alter the moral structure that governs it.[14]

Much of the current debate focuses on the risk of an "accountability vacuum," suggesting that responsibility erodes as systems grow more autonomous and opaque.[15] This chapter argues the opposite. Any accountability gap created by AI is not inherent to the technology; it is a result of deliberate or negligent institutional choice. The actual danger lies not only in algorithmic errors or limited transparency, but in the temptation to treat machine outputs as moral insulation. By interposing layers of computation, confidence scores, and procedural validation between the decision-maker and the use of lethal force, AI can function as a kind of moral car wash—allowing leaders to experience violence as technically authorized rather than personally owned.[16]

The response to the challenge is not to develop ethical frameworks for machines, not to treat legal compliance as a substitute for judgment, and not to prohibit military AI altogether. It is to reassert, deliberately and explicitly, the inalienable responsibilities of human command.[17] Throughout this chapter, responsibility refers not merely to legal liability, but to the commander's obligation to exercise judgment and remain accountable for decisions to use force. Leaders must resist the impulse to treat AI as a substitute for conscience and instead confront the enduring reality that decisions to harm, even when justified, carry an unavoidable moral remainder. As Michael Walzer argued in his account of "dirty hands," commanders may authorize violence to prevent greater evil, but they cannot escape responsibility for the moral remainder that follows.[18] The central challenge of machine-age warfare is ensuring that this burden is neither diluted nor

displaced—that it remains firmly where it belongs, with the humans who decide when and how force is used.

This dilemma cuts both ways. If delegating lethal decision-making to machines risks abdication of responsibility, refusing to employ AI, where it could reasonably reduce civilian harm or protect friendly forces, may also constitute a form of dirty hands. A commander who knowingly selects a more predictable and reliable, but slower or less discriminating human process—when more precise, rigorously constrained but less proven AI-enabled options are available—does not preserve moral innocence by abstention. They accept foreseeable and avoidable harm for the sake of moral safety. In such cases, restraint itself becomes ethically costly. The inescapable tragedy of command persists in AI warfare. Leaders do not get a choice between clean and unclean options because every available path results in harms that carry moral residue.

Clearing the Fog: Distinguishing Ethics from Compliance

Understanding the moral challenges of machine-age warfare requires clearing away a basic but persistent confusion in contemporary debate. Discussions of military AI frequently blur the line between technical performance and ethical judgment.[19] Systems that function "correctly," in the sense that they operate as designed, are often assumed to be acting "ethically."[20] This assumption is false—and dangerous. Ethical conduct in war cannot be reduced to technical compliance. Preserving responsible command judgment requires a clear distinction between adherence to rules and the exercise of judgment, between procedural correctness and responsibility of conscience.

Traceability is not Accountability

One of the most common sources of confusion lies in the concept of traceability.[21] Traceability is a technical attribute that refers to the ability to reconstruct how a system reached a particular output through logs, data provenance, and decision pathways.[22] It is a forensic record. Accountability, by contrast, is a moral and legal condition.[23] It assigns responsibility to a human being who must answer for the consequences of their actions or inactions. The two are related, but they are not interchangeable.

Much of this anxiety surrounding AI centers on the so-called "black box" problem—that opaque, stochastic algorithms make it exceedingly difficult to determine who is responsible when things go wrong. Yet this critique overlooks the fundamental reality of warfare. Human decision-making under stress has always been opaque, shaped by imperfect information, memory, fear, and time pressure. In many cases, and far more objectively than humans, AI-enabled systems can actually enhance traceability by creating detailed records of the information considered and the options rejected.[24] Rather than erasing accountability, such systems can sharpen it. An audit trail does not absolve a commander of responsibility; it provides the evidence required to hold them to account.

The ethical problem posed by AI is not that machines lack responsibility, but that their behavior may be insufficiently predictable.[25] International humanitarian law does not require weapons to possess intent; it requires commanders to reasonably foresee the effects of their use.[26] A system that cannot be reliably understood in a given operational context cannot be ethically or legally employed, regardless of its sophistication. The chain of command

authorizing such use remains morally culpable because it disregarded the obligation to minimize harm in war. In this sense, a simpler and more predictable system may be ethically preferable to a more advanced but inscrutable one. The ethical mandate is not intelligence for its own sake, but control grounded in understanding.[27]

Reliability is Not Moral Judgment

A similar confusion arises around reliability. Reliability, in engineering terms, refers to a system's ability to perform consistently according to its specifications. It does not mean that the system is ethically correct.[28] A system can function exactly as designed and still produce harmful or unjustified outcomes if its design assumptions no longer match reality. This is particularly true when training data diverges from operational conditions. An algorithm trained on urban environments may reliably flag certain behaviors as hostile. Deployed in a rural context, those same behaviors may indicate nothing more than ordinary civilian activity. In such cases, the machine has not failed; the human who relied on its certification without contextual judgment has.

This "compliance trap" is not a new phenomenon.[29] Militaries have long grappled with the dangers of confusing adherence to procedures with responsible judgment.[30] The familiar defense of "following orders" or "following procedure" has been associated with some of the gravest ethical failures in modern history.[31] AI amplifies this risk by automating procedural validation, making it easier for decision-makers to mistake consistency for truth and certification for conscience.

Intent and Responsibility Remain Human

The distinction between ethics and compliance becomes more strained when debates turn to intent. Critics often argue that because AI systems cannot form intent—they calculate probabilities rather than purposes—the ethical framework used to justify foreseen but unintended harm to civilians collapses.[32] If a machine cannot intend, the argument goes, then the logic that permits collateral damage under conditions of necessity and proportionality no longer applies. This objection misunderstands where intent has always resided in war.

Weapons have never been required to possess intent. A rifle does not intend to kill. A cruise missile does not intend to destroy a building. Neither does an algorithm. Intent has always rested with human beings who authorize and employ the use of force.[33] The ethical structure that permits foreseen but unintended harm does not depend on the mental state of the weapon, but on the judgment of the commander who decides that the harm is proportionate to a legitimate military objective and that no less harmful means are available.

The ethical failure arises not because AI lacks intent, but when commanders treat an AI recommendation as a substitute for their own judgment rather than as an input to it.[34] When an algorithm's output is accepted as a binding directive—because it is fast, authoritative, or technically validated—responsibility is not lost to technology; it is surrendered by leadership. The structure of responsibility remains unchanged. Tools provide means. Humans must still provide purpose. Trusting an algorithm without interrogation is ethically indistinguishable from ordering fires based on an unverified report. The scale and speed of data processing may be unprecedented, but the obligation to exercise judgment is not.

This principle is deeply embedded in military practice. U.S. Navy Regulations make clear that command responsibility is absolute.[35] A commanding officer is responsible for the ship even when not personally present or directly involved. The captain, asleep in her stateroom when a negligent collision occurs, is not relieved because she was absent from the bridge. She is relieved because command is inseparable from responsibility. Authority carries liability. Leaders are not punished for every failure; they are removed when they fail to build organizations capable of acting responsibly without constant supervision.

The same logic applies to AI-enabled warfare. Delegating execution, whether to subordinates, automated systems, or decision-support tools, does not entail delegating responsibility.[36] On the contrary, as systems grow more capable and more complex, the burden on commanders intensifies. They must define intent with greater clarity, understand the limits of the tools they employ, and accept that ethical judgment cannot be automated. Clearing the fog between ethics and compliance is therefore not an academic exercise. It is essential to preserving responsible command judgment in a form of warfare that increasingly rewards speed, abstraction, and distance.

Core Ethical Challenges by Domain

The erosion of human responsibility in war manifests differently across domains; yet a familiar pattern emerges: the temptation to allow technological speed, scale, and abstraction to displace human judgment. Many of the ethical dilemmas associated with AI, autonomy, and related technologies are presented as unprecedented. On closer examination, however, they are best understood as

accelerated versions of command challenges that have shaped modern warfare for more than a century.

AI and Autonomy

The integration of autonomous systems into command and control threatens to compress the Observe–Orient–Decide–Act cycle beyond the threshold of unaided human cognition. Critics warn of "flash war" scenarios in which algorithmic processes shift from decision support to decision making, rendering human commanders as liability-bearing bottlenecks.[37] In this framing, operators risk becoming "moral crumple zones" absorbing legal and ethical responsibility while lacking the practical ability to intervene.[38] Senior leaders, on the other hand, may engage in, or benefit from, moral car washing, in which claims of insufficient knowledge, control, and access to evidence concentrate accountability downward and dilute responsibility upward.[39]

This institutional drift highlights why the current debate over intervention is insufficient. Treating real-time intervention as the sole measure of human responsibility, however, creates a false binary. The United States and several allied militaries increasingly argue that control should not be defined by the ability to interrupt every microsecond decision, but by the degree to which system behavior remains aligned with the commander's intent.[40] Meaningful human control is exercised not only at the moment of execution but in the deliberate design, authorization, and constraint of the system.[41]

Under this intent-based framework, human judgment and responsibility are exercised during the planning and parameter-setting phases. When a commander authorizes an autonomous system to

engage specific target categories within defined parameters, and the system operates within those bounds, the commander has not abdicated responsibility. They have delegated execution. Existing International Humanitarian Law remains applicable because the legal and moral obligation rests with the human who authorized the use of force, regardless of the speed of the system.[42] The ethical risk lies not in machine tempo, but in ambiguity or carelessness in defining intent.

The tension between speed and control is not unique to AI. The telegraph, radio, and the intercontinental ballistic missile all compressed decision timelines, forcing commanders to rely on pre-delegated authority and rapid interpretations of sensor data. AI intensifies this dilemma, but it does not create a new moral category. It magnifies the consequences of weak command judgment.[43]

The central danger, therefore, is not that machines act too quickly, but that commanders become passive. Leaders who treat algorithmic recommendations as mandates because "there was no time to question them" have not been betrayed by technology; they have failed the enduring obligation of command. As human reaction becomes less feasible at execution, responsibility shifts upstream, into the careful design of rules of engagement, operational constraints, and authorization decisions.[44]

Some advocates go further, arguing that machine intelligence may be morally superior to human judgment.[45] They note that humans are susceptible to fear, fatigue, anger, and revenge, factors that have historically contributed to unlawful or unethical violence. Properly designed systems, they argue, can adhere to the laws of armed conflict more consistently than stressed human operators, refusing to engage in prohibited actions regardless of emotion or circumstance.

This argument mistakes consistency for morality. A system may be reliable without being ethically sound. Machines do not seek revenge, but neither can they recognize mercy.[46] They follow rules without understanding context. Even minor deviations between training environments and operational reality can produce outcomes that are technically compliant yet morally indefensible.[47] When consistency is mistaken for correctness, violence risks becoming sterile, bureaucratic, and hollow; legally clean but ethically thin.

The ethical danger of AI is often framed as a machine pulling a trigger. A more immediate risk lies in the machine whispering in a commander's ear. AI-enabled decision support systems, particularly those designed to predict adversary intent or recommend preemptive action, introduce subtler threats to restraint, especially at the level of *jus ad bellum*. Two hazards emerge: misplaced responsibility and predicted permissibility.

First, commanders may come to treat algorithmic outputs not as inputs requiring judgment, but as authoritative determinations. As Toni Erskine observes, this represents abdication rather than delegation.[48] When an AI system predicts an imminent attack with high confidence, leaders may experience the decision to strike as technologically compelled rather than morally chosen. The algorithm becomes a "synthetic scapegoat," allowing responsibility to be psychologically displaced even as authority remains unchanged.[49]

Second, predictive analytics can collapse uncertainty into false precision.[50] Probabilistic forecasts transform multiple possible futures into a single, seemingly objective likelihood.[51] This illusion of certainty lowers the thresholds for preemptive or even preventive action by framing violence as a risk management issue rather than a

moral choice.[52] The uncertainty of war, while often tragic, historically reinforces restraint. AI confidence may erode it.

Remote Warfare

Remote warfare, particularly through unmanned aerial systems, is frequently criticized for emotionally distancing operators from violence—the so-called "PlayStation effect."[53] Empirical research and this author's operational experience suggest the opposite.[54] Persistent, high-definition surveillance often produces cognitive combat intimacy rather than detachment, exposing operators to prolonged observation of lethal consequences and moral injury.[55]

The deeper ethical issue posed by remote lethality is not so much emotional detachment but the elimination of the reciprocal risk.[56] Historically, proximity and vulnerability imposed moral friction. Distance removes those constraints, increasing reliance on institutional discipline rather than visceral judgment. Without immediate danger, the act of killing risks being reduced to procedural compliance rather than a grave, necessary choice.

This dynamic raises the risk of ethical deskilling.[57] As Shannon Vallor argues, moral judgment is a cultivated capacity that depends on practice.[58] When ethical discernment is displaced by checklists, interfaces, and algorithmic prompts, the capacity for judgment may atrophy.[59] Remote systems can increase verification time and reduce errors,[60] but absent disciplined leadership, they may also encourage decision-making that is technically correct yet morally anesthetized.

Cyber Operations

Cyber warfare further complicates human responsibility by targeting systems rather than forces. Many cyber operations deliberately affect civilian or dual-use infrastructure, blurring the distinction between combatants and non-combatants. AI magnifies this challenge by enabling operations against the informational and cognitive foundations of civilian life.[61]

Disrupting civilian medical databases, financial systems, and information ecosystems attacks trust itself. AI-enabled cognitive warfare expands propaganda into coercion,[62] treating civilian perception and psychological integrity as military objectives.[63] While states may disagree over whether such actions constitute an armed attack, the ethical implications are clear. Intentionally inducing civilian suffering—whether physical, digital, or psychological—violates the moral logic of distinction regardless of legal classification.[64]

The difficulty lies in attribution, not in ethics. AI-enabled cyber and psychological operations shift the battlespace from physical harm to epistemic and societal harm, but they do not shift moral responsibility away from human decision-makers. The ethical stakes remain centered on how leaders choose to authorize, design, and employ these tools. In this domain, *distinction* becomes epistemic rather than kinetic, as civilian cognition, trust, and perception are often intentionally targeted through human policy choices rather than incidental system effects. *Proportionality* cannot be meaningfully reduced to engagement metrics, sentiment scores, or model confidence, because the most consequential harms—erosion of democratic legitimacy, psychological injury, and degradation of social

trust—are qualitative and cumulative, requiring human judgment rather than technical optimization. *Precaution* is strained by the adaptive and recursive effects of AI, yet uncertainty does not absolve responsibility; it heightens commanders' obligation to anticipate, constrain, and own foreseeable harm rather than treating emergent behavior as exculpatory.

Debates often fixate on technically salient but ethically secondary issues, such as attribution mechanisms, explainability, bias, human-in-the-loop controls, and escalation modeling. These concerns matter, but only instrumentally. They shape the conditions under which responsibility can be exercised, not whether it exists. At the level of *jus ad bellum*, predictive analytics and influence optimization risk reframing necessity as continuous risk management, lowering intervention thresholds while allowing intent to be psychologically displaced onto models and data. The central ethical failure in AI-enabled cognitive warfare is therefore not automation, opacity, or deniability as such, but abdication of judgment—when human authorities allow technical processes to stand in for ethical judgment and when responsibility for manipulating civilian cognition is diffused rather than consciously borne by those who authorize its use.

Biotechnology and the Soldier

Emerging biotechnologies introduce ethical risks of a different kind. Enhancements ranging from stimulants to neural interfaces raise concerns about consent, reversibility, and the authenticity of the enhancements themselves. Critics argue that meaningful consent is already compromised within hierarchical institutions, so

biotechnologies intensify the treatment of soldiers as instruments rather than persons.[65]

Warfare has always sought to enhance lethality, and military service has always involved the instrumentalization of the body.[66] What distinguishes biotechnological enhancement is not the increased reach, speed, firepower, or coercion but the possibility of permanently altering human capacities that support judgment and responsibility. Interventions that suppress fear, hesitation, or remorse may improve performance, but they can also erode the faculties that underpin responsibility.[67] Force may still comply with rules of engagement, but these emotional disconnects remove psychological constraints that risk converting proportionality from judgment to mechanical execution. Technologies that aim to eliminate guilt, hesitation, or remorse also risk eliminating the moral friction that prevents atrocity, rendering violence experientially cheap even when legally justified.

When a system removes the ability to recognize, hesitate before, and take responsibility for harming others, as well as to reflect morally, the individual ceases to function as a fully responsible agent and becomes instrumentalized as a weapons platform.[68] Unlike AI, which threatens to externalize responsibility into systems, biotechnology threatens to internalize command intent directly into human bodies, bypassing reflection. In such cases, responsibility does not disappear; it shifts upward. The chain of command, and ultimately the state, assumes full liability for actions taken by individuals whose capacity for responsible judgment has been deliberately diminished.

The Dirty Hands Moral Remainder

The central ethical danger of algorithmic warfare is not that machines will make the wrong decisions, but that they will enable human beings to make lethal decisions while feeling both legally compliant and morally exonerated. In classical political ethics, the problem of dirty hands refers to the inescapable burden of leadership: circumstances in which harm may be necessary to prevent greater harm yet never becomes morally clean.[69] Authorizing violence against another human being may be justified, but it is never without moral cost.

The defining characteristic of ethical leadership in war is therefore not moral purity but the acknowledgment of this remainder. Guilt, hesitation, and tragic awareness are not failures of character; they are indicators of responsibility taken seriously. This moral friction restrains violence and prevents necessity from sliding into indifference. The integration of AI threatens to erode that restraint by making it easier for those with authority to distance themselves from the consequences of their decisions. A sanitized conscience through automation may feel efficient, but "unreflective adoption of [such] tools" hollows out the ethical core of the profession and risks normalizing violence.[70] These domain-specific challenges converge on a deeper problem—one that technology does not resolve but threatens to conceal. As Martin Luther King Jr. warned, "When scientific power outruns moral power, we end up with guided missiles and misguided men."[71] The danger in machine-age warfare is not only misguided machines, but misguided leaders who mistake technical process for moral judgment.

Algorithmic Fragmentation and Moral Dissociation

The ethical risk of military AI lies in its capacity to fragment responsibility. Lethal decisions are divided into micro-processes—data ingestion, probabilistic classification, confidence scoring, and recommendation—allowing each participant to view themselves as a technical contributor rather than a responsible actor.[72] Responsibility becomes diffused across systems and roles, encouraging dissociation rather than ownership.[73]

Hannah Arendt identified a similar pattern in mid-twentieth-century bureaucratic systems, in which responsibility was diluted through administrative processes.[74] Contemporary observers note that technological complexity can produce a comparable effect, transforming warriors into compliance officers.[75] AI intensifies this risk by combining speed, abstraction, and opacity. Advanced targeting systems now streamline collateral assessments and validation checks, making it easier to reduce ethical deliberation to interface confirmation rather than substantive judgment.[76]

The danger is not efficiency itself, but passivity. When lethal authorization becomes contingent on a "green light" produced by a confidence score, the moral remainder is not resolved; it is systematically obscured. Decision-makers are tempted to rationalize. "I did not choose to kill; the system certified the risk as acceptable." Violence becomes anonymous and administratively sanitized. The moral restraint that once constrained excess begins to erode when responsibility is displaced onto software.

The deeper tension in AI-enabled targeting lies between computational precision and human judgment. Algorithms can

calculate blast radii and casualty estimates, but proportionality is not a mathematical output. It requires qualitative judgment about meaning, context, and consequences. An algorithm may quantify physical damage to a religious site, but it cannot evaluate cultural, political, or spiritual injury.[77] A commander who relies solely on numerical estimates without engaging those qualitative dimensions is not exercising judgment; they are seeking distance from responsibility.

Proportionality cannot be encoded in any complete or morally sufficient way. It requires a human decision-maker willing to weigh incommensurable values and accept the consequences of that choice. Attempts to replace judgment with calculation do not eliminate moral risk; they convert it into negligence, willful blindness, or moral injury. Abdication is not an option. The burden remains human.

Double Effect and the Intentionality Gap

The increasing automation of the targeting cycle strains the ethical framework that has traditionally justified unintended harm to civilians. The doctrine of double effect permits foreseen but unintended harm only when the military objective is necessary, proportional, and intended.[78] Critics argue that because algorithms cannot form intent, the framework collapses.[79]

The deeper fracture, however, is not technological but human. When commanders defer to algorithmic target selection without understanding why a system identified a target, they substitute statistical confidence for human intention and judgment.[80] What was once a tragic decision becomes a technical outcome. Civilian harm is attributed to model error rather than command choice, creating an

intentionality gap in which leaders claim strategic benefit while disowning moral cost.

Others counter, correctly, that the doctrine of double effect does not require weapons to possess intent.[81] Rifles, missiles, and artillery never formed intent; responsibility always rested with those who employed them. AI does not change this structure. Intent remains with the human who authorizes the use of force.[82]

The ethical crisis arises when intent is poorly specified or psychologically displaced. For AI to function as an extension of command rather than a substitute for conscience, commanders must articulate intent with greater precision than ever before. The requirement is not perfect prediction, but deliberate ownership. When intent is clear, accountability remains intact. When it is vague, technology becomes an excuse to evade responsibility.

Risk Reduction and the Limits of Sanitization

Arguments in favor of autonomous systems frequently emphasize risk reduction. If machines can lessen physical exposure or psychological strain on friendly forces, some proponents suggest that commanders have a strong ethical reason, if not an obligation, to employ them.[83] This logic increasingly shapes defense policy, treating force protection and reduced operator risk as legitimate objectives of autonomy.[84] The argument is compelling, but incomplete.

Moral injury is not merely a harm to be minimized;[85] it is also a signal that violence retains its gravity. The psychological burden associated with killing has historically imposed restraint, reinforcing the awareness that the use of force is a tragic necessity rather than a technical task.[86] Technologies that promise frictionless lethality risk

converting war into a logistical exercise rather than a profoundly human undertaking.[87]

A force capable of killing without hesitation may be operationally efficient, but it also risks becoming ethically hollow. The danger is not the reduction of suffering per se, but the erosion of moral awareness. When lethal action no longer carries weight—when it is experienced as clean, distant, or routine, the moral friction that once constrained excess begins to erode.

Tragic Judgment and Human Dignity

AI systems foster the belief that violence can be calibrated without moral risk—that sufficient data can eliminate moral uncertainty. Resisting the illusion requires what has been described as tragic judgment: the recognition that harm remains harm even when justified.[88] Doubt is not a weakness. It is moral humility.

The authority to kill must remain human, not only for reasons of accountability, but for reasons of dignity. War presupposes an interpersonal relationship between adversaries. Even at its most brutal, it acknowledges shared humanity. To be killed by a human is tragic. To be processed and terminated by an algorithm is, many argue, an indignity. It reduces the enemy to a data point, whose removal from the correlation of forces becomes an administrative procedure.[89] From a deontological perspective concerned with human dignity, this constitutes a moral violation: the ethical quality of the act is shaped in part by the reciprocal human-to-human relationship it presupposes.

Consequentialist approaches, by contrast, which judge moral permissibility primarily by outcomes, often contest whether this distinction matters. They argue that if an algorithm produces more

discriminating and less cruel outcomes than a human, the perceived indignity may be outweighed by the preservation of life.[90] Nonetheless, the intuition remains central to the ethics of human dignity in war.

Preserving human responsibility in machine-age warfare, therefore, requires preserving risk and friction—psychological, ethical, and institutional. The burden of dirty hands cannot be delegated. It must be borne.

Command Responsibility in a Tech-Centric Force

If human responsibility is to be preserved in machine-age warfare, it cannot rely solely on individual virtue; it must be reinforced through leadership, education, and institutional design. This is the central leadership challenge of the technological age: ensuring that operational advantage does not come at the cost of human judgment and responsibility. History suggests that wars are not lost because militaries lack advanced tools, but because they fail to exercise political judgment and govern their use responsibly.[91]

Ronald Reagan went further, arguing that this moral foundation is America's greatest strength against adversaries who view ethical restraint as a weakness: "Above all, we must realize that no arsenal, or no weapon in the arsenals of the world, is so formidable as the will and moral courage of free men and women. It is a weapon our adversaries in today's world do not have."[92] Whatever one's view of Reagan's politics, the underlying claim that moral courage is a strategic asset rather than a liability remains salient today. In an era of rapid technical parity, this moral foundation is a key differentiator in today's Great Power Competition, and the American warrior ethos is

a force multiplier for defending the homeland and deterring adversaries.

The task for senior leaders, therefore, is not to invent new ethical codes, but to translate enduring obligations of command into a technological environment defined by speed, abstraction, and delegation. Legal compliance alone is insufficient. Human-in-the-loop requirements are meaningless if human judgment is reduced to procedural approval rather than substantive decision-making. Preserving responsibility in machine-age warfare requires deliberate leadership choices across education, systems design, and accountability structures. Three mutually reinforcing requirements are essential: epistemic competence, structural friction, and responsibility by design.

From Compliance to Competence

Contemporary military ethics education is often framed primarily in terms of compliance with rules and regulations rather than the cultivation of judgment under conditions of moral stress.[93] While mastery of the law of armed conflict remains essential, compliance alone does not prepare commanders to exercise judgment in AI-enabled operations. Algorithmic systems introduce new forms of uncertainty—statistical inference, probabilistic confidence, and data dependency—that demand a deeper form of professional competence.

Adherence to rules often fails to address the "hard cases" or novel complexities of AI-enabled warfare, so military education must pivot toward a virtue-ethics framework that emphasizes *phronesis*—the character and practical wisdom required to apply general rules to specific contexts.[94] Unlike static algorithms or legal checklists,

phronesis enables soldiers to interpret general rules with contextual sensitivity, balancing competing values such as courage and mercy to meet the specific moral demands of a situation.[95]

Training and education should focus on preventing "moral deskilling" through repeated exercise of judgment in morally contested cases or dilemmas in which rules conflict or no clear "right" answer exists.[96] Such cultivation of character also relies on human role models who demonstrate how to apply lethal force with integrity. This is not a rejection of rules-based ethics or the law of armed conflict, but a recognition that no rule set can anticipate every morally salient contingency introduced by AI-enabled operations.

Commanders must understand not only what a system recommends, but how and why it produces that recommendation. Before authorizing AI-enabled operations, leaders must meet a standard of epistemic competence: a level of understanding sufficient to calibrate trust appropriately.[97] This includes awareness of training data limitations, operational blind spots, failure modes, and the potential for model error or drift. A commander who cannot explain when, why, and under what operational conditions a system's output is reliable—or less likely to be so, has not earned the authority to delegate lethal action to it.

Equally important is organizational clarity. Opacity cannot become an excuse for evading responsibility. Leaders and operators must understand that "the model made me do it" is not a defensible justification. Trusting an algorithm without interrogation is not merely a technical mistake; it is a failure of command. In a profession where authority and responsibility are inseparable, unexamined ignorance of the system one authorizes is disqualifying. Ethical employment of

advanced systems requires professional literacy commensurate with the risks they pose.

Designing Friction, Not Illusion of Control

Calls for a physical "stop button" on autonomous systems reflect a desire for reassurance, but they risk misunderstanding the nature of control in high-tempo operations. In environments shaped by hypersonic speeds, cyber effects, and swarming systems, real-time human intervention may be impractical or counterproductive. Control cannot be reduced to a momentary override.

A more credible approach is to design structural friction into systems and institutions. Meaningful human control is exercised upstream through rigorous testing, operational constraints, legal review, and clear authorization frameworks.[98] Control is maintained not by interrupting execution at the last second, but by shaping what systems are permitted to do in the first place.

This ensures the ethical burden remains squarely on leadership decisions made before employment. Commanders are responsible for ensuring that systems are designed and configured without sufficient confidence that they will not produce unlawful or unethical effects. Responsibility does not dissipate once a weapon is released; it reflects the quality of judgment exercised at the point of authorization. As with any weapon, control lies in the decision to activate the system and the conditions under which it is employed, not in the fantasy of stopping it mid-flight.

This concern is not limited to Western debates. China's position paper submitted to the Convention on Certain Conventional Weapons explicitly affirms that responsibility for military applications of AI

must remain with human actors, underscoring that across competing strategic cultures, accountability in AI-enabled warfare is understood as a matter of governance and command responsibility rather than machine autonomy.[99]

Accountability by Design

Finally, accountability must be embedded institutionally rather than asserted rhetorically. Responsibility by design requires that systems be deployed only when clear lines of human accountability can be demonstrated.[100] This includes transparent system architecture, documented testing and validation, and operational rules that preserve traceability throughout the weapon's life cycle.[101]

Technical measures such as tamper-resistant logging and decision-path recording support accountability, but they do not replace it. Their purpose is to ensure that every lethal outcome can be traced to human authorization decisions. Accountability must never be anonymous; responsibility must never be displaced. Risk acceptance, design tradeoffs, and operational shortcuts must be attributable to identifiable actors.

A distributed model of responsibility clarifies rather than dilutes accountability. Developers bear responsibility for foreseeable design limitations and biases. Operators are responsible for making contextual judgments and employing systems appropriately. Commanders retain ultimate responsibility for authorizing employment.[102] States are accountable for the legal and institutional frameworks that govern these systems.[103] This division does not diffuse moral burden; it assigns it where authority resides.

Conclusion: The Unchanging Burden of Command

AI and related technologies are transforming the *character* of war—its speed, reach, and modes of execution. They do not, however, alter the *nature* of war itself. Armed conflict remains a human enterprise, shaped by political purpose, uncertainty, and the deliberate use of violence to pursue strategic ends. The enduring ethical challenge, therefore, is not how machines behave, but how humans choose to employ them.

Much of the contemporary anxiety surrounding "killer robots" distracts from a more immediate and consequential risk:[104] the quiet evasion of moral responsibility through technological delegation. The gravest danger is not that machines will seize control, but that commanders and institutions will surrender judgment in exchange for speed, certainty, or procedural cover. When lethal force is filtered through layers of computation and abstraction, the temptation to treat outcomes as technologically determined rather than morally chosen becomes acute.

No technology can resolve the moral tragedy inherent in war. No algorithm can sanitize the decision to take human life, nor any system that can absorb that remainder on behalf of those who authorize it. Attempts to shift blame to software, models, or opaque systems do not eliminate responsibility; they merely obscure it. The burden of command—the necessity to decide, to accept risk, and to bear the consequences—remains irreducibly human.

Preserving responsibility in machine-age warfare, therefore, requires a reaffirmation of the profession's most fundamental principle: responsibility follows authority. This principle is the

foundation of trust upon which the professionalism and cohesion of the U.S. military are built. As systems grow more autonomous and more capable, the obligations of command do not diminish; they intensify, because authority is exercised earlier, farther upstream, and with wider consequences. The strength and lethality of America's fighting forces, which have effectively deterred adversary aggression, depend upon leaders ensuring that advanced technologies remain instruments of human judgment even as execution becomes distributed and automated.

This chapter argues that ethical leadership in a technologically advanced force does not depend on new moral theories or novel legal constructs. It depends on disciplined command practice: leaders who understand the systems they employ, who define intent with clarity, who design structures that preserve accountability, and who accept that no amount of technical sophistication can relieve them from moral ownership. In the end, technological advantage will matter, but only if it is governed by leaders willing to carry the full weight of the decisions they make.

Machines may increasingly shape the future of warfare. Responsibility for its conduct, however, cannot be delegated.

Endnotes

1 The author is solely responsible for the views expressed in this publication, which do not necessarily represent the official policy or position of the Daniel K. Inouye Asia-Pacific Center for Security Studies, the U.S. Department of War, or the U.S. government.

2 U.S. Congress, Joint Committee on Atomic Energy, *Radiation Safety and Regulation: Hearings before the Joint Committee on Atomic Energy*, 87th Cong., 1st sess., June 12–15, 1961 (Washington, DC: Government Printing

Office, 1961), 366, https://exhibits.stanford.edu/atomic-energy/catalog/jr755wm3415.

[3] Bessma Momani, Aaron Shull, and Jean-François Bélanger, "Introduction: The Ethics of Automated Warfare and AI*," Centre for International Governance Innovation*, November 28, 2022, https://www.cigionline.org/articles/introduction-the-ethics-of-automated-warfare-and-ai/; Thilo Hagendorff, "Mapping the Ethics of Generative AI: A Comprehensive Scoping Review," *Minds and Machines* 34. No. 39 (2024), https://doi.org/10.1007/s11023-024-09694-w; Toby Walsh, "The Problem with Artificial (General) Intelligence in Warfare," *Centre for International Governance Innovation*, November 28, 2022, https://www.cigionline.org/articles/the-problem-with-artificial-general-intelligence-in-warfare/.

[4] Momani, Shull, and Bélanger, "Introduction: The Ethics of Automated Warfare and AI"; Hagendorff, "Mapping the Ethics of Generative AI."

[5] James Johnson, "The AI Commander Problem: Ethical, Political, and Psychological Dilemmas of Human-Machine Interactions in AI-Enabled Warfare," *Journal of Military Ethics* 21, no. 3–4 (2022): 246–71, https://doi.org/10.1080/15027570.2023.2175887; Alessandro Nalin and Paolo Tripodi, "Future Warfare and Responsibility Management in the AI-Based Military Decision-Making Process," *Journal of Advanced Military Studies* 14, no. 1 (2023): 83–97, https://doi.org/10.21140/mcuj.20231401003; Neil Renic, "Tragic Reflection, Political Wisdom, and the Future of Algorithmic War," *Australian Journal of International Affairs* 78, no. 2 (2024): 247–56, https://doi.org/10.1080/10357718.2024.2328299.

[6] Toni Erskine, "Before Algorithmic Armageddon: Anticipating Immediate Risks to Restraint When AI Infiltrates Decisions to Wage War," *Australian Journal of International Affairs* 78, no. 2 (2024): 175–90, https://doi.org/10.1080/10357718.2024.2345636; Davit Khachatryan, "Military AI Challenges Human Accountability," *Center for International Policy*, July 21, 2025, https://internationalpolicy.org/publications/military-ai-challenges-human-accountability/.

[7] Zoe Stanley-Lockman, *Responsible and Ethical Military AI: Allies and Allied Perspectives*, CSET Issue Brief (Washington, DC: Center for Security and Emerging Technology, 2021), https://cset.georgetown.edu/wp-content/uploads/CSET-Responsible-and-Ethical-Military-AI.pdf; Jinghan Zeng, "The US Factor in Chinese Perceptions of Militarized Artificial Intelligence," *International Affairs* 101, no. 2 (2025): 677–689, https://doi.org/10.1093/ia/iiae323.

[8] Elsa B. Kania, *Battlefield Singularity: Artificial Intelligence, Military Revolution, and China's Future Military Power* (Washington, DC: Center for a New American Security, 2017), https://www.cnas.org/publications/reports/battlefield-singularity-artificial-intelligence-military-revolution-and-chinas-future-military-power; John Lee, "'Overtaking on the Curve'? Defense AI in China," in *The Very Long Game: 25 Case Studies on the Global State of Defense AI*, ed. Heiko Borchert et al. (Cham, Switzerland: Springer, 2024), 465–86, https://doi.org/10.1007/978-3-031-58649-1_21; Kurtis H. Simpson et al., "Militarizing AI: How to Catch the Digital Dragon?" *Centre for International Governance Innovation*, February 26, 2025, https://www.cigionline.org/articles/militarizing-ai-how-to-catch-the-digital-dragon/; Jacob Stokes, "Military Artificial Intelligence, the People's Liberation Army, and U.S.-China Strategic Competition," testimony before the U.S.–China Economic and Security Review Commission, Hearing on *Current and Emerging Technologies in U.S.–China Economic and National Security Competition*, February 1, 2024, https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/Jacob_Stokes_Testimony.pdf.

[9] Thomas X. Hammes, "Autonomous Weapons Are the Moral Choice," *New Atlanticist* (blog), Atlantic Council, November 2, 2023, https://www.atlanticcouncil.org/blogs/new-atlanticist/autonomous-weapons-are-the-moral-choice/; Lauren A. Kahn, "Risky Incrementalism: Defense AI in the United States," in *The Very Long Game; 25 Case Studies on the Global State of Defense AI*, ed. Heiko Borchert et al. (Cham, Switzerland: Springer), 39–61, https://link.springer.com/chapter/10.1007/978-3-031-58649-1_2; Rosario M. Simonetti, "Command and Control in the Age of Automated Decision Making Process, Challenges and Opportunities" (master's thesis, Marine Corps University, 2019), https://apps.dtic.mil/sti/trecms/pdf/AD1177394.pdf; Alex Wilner, "AI and the Future of Deterrence: Promises and Pitfalls," *Centre for International Governance Innovation*, November 28, 2022, https://www.cigionline.org/articles/ai-and-the-future-of-deterrence-promises-and-pitfalls/.

[10] Hammes, "Autonomous Weapons Are the Moral Choice"; Kahn, "Risky Incrementalism."

[11] John R. Allen and Amir Husain, "On Hyper War," *Proceedings* 143, no 7 (2017), https://www.usni.org/magazines/proceedings/2017/july/hyperwar.

[12] Antulio J. Echevarria II, "War's Changing Character and Varying Nature: A Closer Look at Clausewitz's Trinity," *Parameters* 5, no. 4 (2017), https://www.militarystrategymagazine.com/article/wars-changing-character-and-varying-nature-a-closer-look-at-clausewitzs-trinity/.

[13] Khachatryan, "Military AI Challenges"; Seth D. Norrholm et al., "Remote Warfare with Intimate Consequences: Psychological Stress in Service Member and Veteran Remotely-Piloted Aircraft (RPA) Personnel," *Journal of Mental Health & Clinical Psychology* 7, no. 3 (2023): 37–49, https://doi.org/10.29245/2578-2959/2023/3.1289; Erich Riesen, "The Moral Case for the Development and Use of Autonomous Weapon Systems," *Journal of Military Ethics* 21, no. 2 (2022): 132–50, https://doi.org/10.1080/15027570.2022.2124022.

[14] Hammes, "Autonomous Weapons"; Zeng, "The US Factor."

[15] Rebecca Crootof, "AI and the Actual IHL Accountability Gap," *Centre for International Governance Innovation,* November 28, 2022, https://www.cigionline.org/articles/ai-and-the-actual-ihl-accountability-gap/; Jie Guo, "The Ethical Legitimacy of Autonomous Weapons Systems: Reconfiguring War Accountability in the Age of Artificial Intelligence," *Ethics & Global Politics* 18, no. 3 (2025): 27–39, https://doi.org/10.1080/16544951.2025.2540131.

[16] Massimiliano Lorenzo Cappuccio et al., "A Taste of Armageddon: A Virtue Ethics Perspective on Autonomous Weapons and Moral Injury," *Journal of Military Ethics* 21, no. 1 (2022): 19–38, https://doi.org/10.1080/15027570.2022.2063103; Crootof, "AI and the Actual IHL Accountability Gap"; Erskine, "Before Algorithmic Armageddon"; Johnson, "The AI Commander Problem"; Shaun Waterman, "Can AI Help Targeteers Curb Civilian Casualties?" *Air & Space Forces Magazine*, August 11, 2025, https://www.airandspaceforces.com/can-ai-help-targeteers-curb-civilian-casualties/.

[17] Khachatryan, "Military AI Challenges."

[18] Renic, "Tragic Reflection"; Michael Walzer, "Political Action: The Problem of Dirty Hands," *Philosophy & Public Affairs* 2, no. 2 (1973): 161–62, https://www.jstor.org/stable/2265139.

[19] Laura Bruun and Marta Bo, *Bias in Military Artificial Intelligence and Compliance with International Humanitarian Law* (Stockholm: SIPRI, 2025), https://doi.org/10.55163/NLWV5347; Guo, "The Ethical Legitimacy"; David Oniani et al., "Adopting and Expanding Ethical Principles for Generative Artificial Intelligence from Military to Healthcare," *NPJ Digital Medicine* 6, no. 1 (2023): 225, https://doi.org/10.1038/s41746-023-00965-x.

[20] Bruun and Bo, *Bias in Military AI*.

[21] Oniani et al., "Adopting and Expanding."

[22] Oniani et al., "Adopting and Expanding."

[23] Mark Bovens, "Analysing and Assessing Accountability: A Conceptual Framework," *European Law Journal* 13, no 4. (2007), 447–68, https://doi.org/10.1111/j.1468-0386.2007.00378.x.

[24] Mariarosaria Taddeo et al., "Ethical Principles for Artificial Intelligence in National Defence," *Philosophy & Technology* 34, no. 4 (2021): 1707–29, https://doi.org/10.1007/s13347-021-00482-3.

[25] Laura Bruun, Marta Bo, and Netta Goussac, *Compliance with International Humanitarian Law in the Development and Use of Autonomous Weapon Systems* (Stockholm: SIPRI, 2023), https://www.sipri.org/sites/default/files/2023-03/ihl_and_aws.pdf; Amanda M. Eklund, *Meaningful Human Control of Autonomous Weapon Systems: Definitions and Key Elements in the Light of International Humanitarian Law and International Human Rights Law* (Stockholm: Swedish Defence Research Agency [FOI], 2020), https://www.fcas-forum.eu/publications/Meaningful-Human-Control-of-Autonomous-Weapon-Systems-Eklund.pdf; Xiuwei Li. "Lei Ruipeng: Rengong zhineng de lunli xue he zhili" [Lei Ruipeng: Ethics and Governance of Artificial Intelligence], *Zhongguo Shehui Kexue Wang* [China Social Sciences Net], November 24, 2023, https://www.cssn.cn/zx/zx_rdkx/202311/t20231124_5698790.shtml.

[26] Protocol Additional to the Geneva Convention of 12 August 1949, and Relating to the Protection of Victims of International Armed Conflicts (Protocol I), June 8, 1977, 1125 United Nations Treaty Series 3, arts 51(5)(b), 52(2), https://treaties.un.org/doc/Publication/UNTS/Volume%201125/volume-1125-I-17512-English.pdf; Yoram Dinstein, *The Conduct of Hostilities Under the Law of International Armed Conflict*, 3rd ed. (Cambridge: Cambridge University Press, 2016), 155–58.

[27] Lei, "Ethics and Governance."

[28] Oniani et al., "Adopting and Expanding."

[29] Alexander P. Holder, "The Compliance Trap and the Protection of Civilians: An Empirical Investigation of 'The Achievement of Legality' in an Aerial Targeting Operation," Article 36, December 2021, compliance-trap-alexander-holder-singlepage.pdf.

[30] Guo, "The Ethical Legitimacy"; Conrad Sanderson, David Douglas, and Qinghua Lu, "Implementing Responsible AI: Tensions and Trade-Offs Between Ethics Aspects," in *Proceedings of the International Joint Conference on Neural Networks*, *(IJCNN)* (Gold Coast, Australia, 2023), https://doi.org/10.1109/IJCNN54540.2023.10191274; Shannon Vallor, "The Future of Military Virtue: Autonomous Systems and the Moral Deskilling of the Military," in *2013 5th International Conference on Cyber Conflict (CyCon*

2013), ed. Karlis Podens, Jan Stinissen, and Markus Maybaum (Tallinn, Estonia: NATO Cooperative Cyber Defense Centre of Excellence, 2013), 471–86, https://scholarcommons.scu.edu/cgi/viewcontent.cgi?article=1001&context=phi.

31 Sanderson et al., "Implementing Responsible AI."

32 Guo, "The Ethical Legitimacy"; Khachatryan, "Military AI Challenges"; "Nonmaleficence, Edition 1.0 Research," in *AI & Human Rights Index*, ed. Nathan C. Walker et al., (New York: All Tech is Human; Camden, NJ: AI Ethics Lab at Rutgers University, 2026), https://aiethicslab.rutgers.edu/glossary/nonmaleficence/.

33 Bruun and Bo, *Bias in Military AI*; Guo, "The Ethical Legitimacy"; Khachatryan, "Military AI Challenges."

34 Bruun and Bo, *Bias in Military AI*; Johnson, "The AI Commander Problem."

35 U.S. Department of the Navy, U.S. Navy Regulations, 1990, Chapter 8, "The Commanding Officer" (Washington, DC, 1990), 47, https://www.secnav.navy.mil/doni/US%20Navy%20Regulations/Chapter%208%20-%20The%20Commanding%20Officer.pdf.

36 Kraska, James. "Command Accountability for AI Weapon Systems in the Law of Armed Conflict." International Law Studies 97, no. 1 (2021): 22.

37 Guo, "The Ethical Legitimacy"; Nalin and Tripodi, "Future Warfare"; Walsh, "Problem with A(G)I in War."

38 Madeleine Clare Elish. "Moral Crumple Zones: Cautionary Tales in Human-Robot Interaction," *Engaging Science, Technology, and Society* 5 (2019): 40–60, https://doi.org/10.17351/ests2019.260; *AI & Human Rights Index*.

39 Nalin and Tripodi, "Future Warfare"; Robert Sparrow, "Robots and Respect: Assessing the Case Against Autonomous Weapon Systems," *Ethics & International Affairs* 30, no. 1 (2016): 93–116, https://doi.org/10.1017/S0892679415000647.

40 U.S. Department of Defense, *Directive 3000.09, Autonomy in Weapon Systems* (Washington, DC: DOD, January 25, 2023), sec. 4, https://www.esd.whs.mil/Portals/54/Documents/DD/issuances/dodd/300009p.pdf; Kearney, "Preserving Human Moral Agency as Use of AI-Driven Autonomous Weapons Grows," *Kearney Insights*, May 15, 2025, https://www.kearney.com/industry/aerospace-defense/paris-air-show/article/preserving-human-moral-agency-as-use-of-ai-driven-autonomous-weapons-grows.

[41] Eklund, "Meaningful Human Control"; Taddeo et al., "Ethical Principles"; Filippo Santoni de Sio and Jeroen van den Hoven, "Meaningful Human Control over Autonomous Systems: A Philosophical Account," *Frontiers in Robotics and AI* 5 (2018): 15, https://doi.org/10.3389/frobt.2018.00015; Masoud Zamani, "How Meaningful is 'Meaningful Human Control' in LAWS Regulation?" *Lieber Institute*, U.S. Military Academy at West Point, March 26, 2025, https://lieber.westpoint.edu/how-meaningful-is-meaningful-human-control-laws-regulation/.

[42] Kraska, "Command Accountability."

[43] Johnson, "The AI Commander Problem"; Nalin and Tripodi, "Future Warfare."

[44] Kraska, "Command Accountability."

[45] Guo, "The Ethical Legitimacy"; Riesen, "The Moral Case."

[46] Bruun and Bo, *Bias in Military AI*; Cappuccio et al., "A Taste of Armageddon."

[47] Bruun and Bo, *Bias in Military AI*.

[48] Erskine, "Before Algorithmic Armageddon."

[49] Erskine, "Before Algorithmic Armageddon."

[50] Erskine, "Before Algorithmic Armageddon," 185.

[51] Erskine, "Before Algorithmic Armageddon," 185.

[52] Erskine, "Before Algorithmic Armageddon," 184; Wilner, "AI and the Future of Deterrence."

[53] Chris Cole, *Convenient Killing: Armed Drones and the 'PlayStation' Mentality,* Fellowship of Reconciliation, September 20, 2010, https://dronewars.net/2010/09/20/convenient-killing-armed-drones-and-the-playstation-mentality-2/; Wayne L. Chappelle et al., "Symptoms of Psychological Distress and Post-Traumatic Stress Disorder in United States Air Force 'Drone' Operators," *Mil Med* 179, no 8 (2014): https://doi.org/10.7205/MILMED-D-13-00501; Seth Norrholm et al., "Remote Warfare with Intimate Consequences," 40–41.

[54] Johnson, "The AI Commander Problem"; Norrholm et al., "Remote Warfare with Intimate Consequences"; Arthur P. Rebera and Nicholas G. Wood, "Virtues and Rules in War: Military Ethics and Technologies of Radical Risk-Reduction," *Ethical Theory and Moral Practice* (2025), https://doi.org/10.1007/s10677-025-10515-x.

[55] Cappuccio et al., "A Taste of Armageddon"; Norrholm et al., "Remote Warfare with Intimate Consequences," 40–41.

[56] Guo, "The Ethical Legitimacy"; Rebera and Wood, "Virtues and Rules."

[57] Vallor, "The Future of Military Virtue."

[58] Vallor, "The Future of Military Virtue," 471; Guo, "The Ethical Legitimacy"; Cappuccio et al., "A Taste of Armageddon."

[59] David Kirichenko, "How Will Artificial Intelligence Impact Battlefield Operations?" *Lawfare*, April 8, 2025, https://www.lawfaremedia.org/article/how-will-artificial-intelligence-impact-battlefield-operations; Vallor, "The Future of Military Virtue."

[60] Rebera and Wood, "Virtues and Rules."

[61] Long Ying [龙瀛], Du Juan [杜鹃], Ma Kechun [马克春], "Junmin ronghe zhanlüe shiyu zhong de rengong zhineng yingyong lunli yanjiu baogao—Zhongwai rengong zhineng yingyong lunli fengxian fenxi" [Research Report on AI Application Ethics from the Perspective of Military-Civilian Integration Strategy: Analysis of Ethical Risks of AI Application at Home and Abroad], *Kepu Zhongguo* [Kepuchina], December 30, 2025, https://cloud.kepuchina.cn/newSearch/imgText?id=7263576697273970688; Simpson et al., "Militarizing AI."

[62] Long, Du, and Ma, "Research Report on AI Application Ethics"; Stokes, "MAI, PLA, and Strategic Competition."

[63] Kania, *Battlefield Singularity*.

[64] Long, Du, and Ma, "Research Report on AI Application Ethics"; Eleonore Pauwels, "Civilian Data in Cyberconflict: Legal and Geostrategic Considerations," *Centre for International Governance Innovation*, November 28, 2022, https://www.cigionline.org/articles/civilian-data-in-cyberconflict-legal-and-geostrategic-considerations/.

[65] Heather A. Harrison Dinniss and Jann K. Kleffner, "Soldier 2.0: Military Human Enhancement and International Law," *International Law Studies* 92 (2016): 432–82, https://digital-commons.usnwc.edu/ils/vol92/iss1/14/.

[66] Harrison Dinniss and Kleffner, "Soldier 2.0."

[67] Harrison Dinniss and Kleffner, "Soldier 2.0."

[68] Cappuccio et al., "A Taste of Armageddon"; Harrison Dinniss and Kleffner, "Soldier 2.0"; Lei, "Ethics and Governance."

[69] Renic, "Tragic Reflection"; Walzer, "Political Action," 161–62, 172–75; Hannah Bleher and Matthias Braun, "Reflections on Putting AI Ethics into

Practice: How AI Ethics Approaches Conceptualize Theory and Practice," *Sci Eng Ethics* 29, no 21 (2023), https://doi.org/10.1007/s11948-023-00443-3.

70 Guo, "The Ethical Legitimacy."

71 King, Martin Luther, Jr. "The Three Evils of Society." Speech, National Conference for New Politics, Chicago, IL, August 31, 1967.

72 Bianca Baggiarini, "Algorithmic War and the Dangers of In-visibility, Anonymity, and Fragmentation," *Australian Journal of International Affairs* 78, no. 2 (2024): 257–65, https://doi.org/10.1080/10357718.2024.2333824; Alexander Blanchard and Laura Bruun, *Autonomous Weapon Systems and AI-enabled Decision Support Systems in Military Targeting: A Comparison and Governance Implications* (Stockholm: SIPRI, 2025), https://doi.org/10.55163/YQBY3151; Taddeo et al., "Ethical Principles."

73 Baggiarini, "Algorithmic War."

74 Hannah Arendt, *Eichmann in Jerusalem: A Report on the Banality of Evil* with an introduction by Amos Elon (New York: Penguin Books, 2006), 135–50, 276–79, 287–88.

75 Crootof, "AI and the Accountability Gap"; U.S. Department of War, "Secretary Hegseth's Message to the Force," January 25, 2025, https://www.war.gov/News/Releases/Release/Article/4040940/secretary-hegseths-message-to-the-force/.

76 Waterman, "Can AI Help Targeteers?"

77 Eklund, "Meaningful Human Control"; Kearney, "Preserving Human Moral Agency"; Renic, "Tragic Reflection."

78 Bruun and Bo, *Bias in Military AI*; *AI & Human Rights Index*.

79 Guo, "The Ethical Legitimacy."

80 Johnson, "The AI Commander Problem."

81 Michael Walzer, *Just and Unjust Wars: A Moral Argument with Historical Illustrations*, 5th ed. (New York: Basic Books, 2015) 152–59; Dinstein, *The Conduct of Hostilities Under the Law of International Armed Conflict*, 131–35; International Committee of the Red Cross, *Commentary on the Additional Protocols of 8 June 1977*, ed. Yves Sandoz, Christophe Swinarski, and Bruno Zimmermann (Geneva: ICRC, 1987), 684–86.

82 Kraska, "Command Accountability."

83 Hammes, "Autonomous Weapons Are the Moral Choice"; Ronald C. Arkin, *Governing Lethal Behavior in Autonomous Robots* (Boca Raton, FL: CRC

Press, 2009), 29–36; Cappuccio et al., "A Taste of Armageddon"; Riesen, "The Moral Case."

84 U.S. DOD, *Directive 3000.09, Autonomy in Weapon Systems*

85 Cappuccio et al., "A Taste of Armageddon."

86 Guo, "The Ethical Legitimacy."

87 Cappuccio et al., "A Taste of Armageddon"; Renic, "Tragic Reflection."

88 Renic, "Tragic Reflection."

89 Rebera and Wood, "Virtues and Rules"; Sparrow, "Robots and Respect."

90 Arkin, *Governing Lethal Behavior in Autonomous Robots*, 37–48. Arkin represents the functionalist or consequentialist school, which prioritizes reducing war crimes over the humanity of the weapon.

91 Renic, "Tragic Reflection"; Simonetti, "Command and Control."

92 Ronald Reagan, "'Above All, We Must Realize That No Arsenal or No Weapon in the Arsenals of the World Is So Formidable as the Will and Moral Courage of Free Men and Women,'" inaugural address, January 20, 1981, Ronald Reagan Presidential Foundation & Institute, https://www.reaganfoundation.org/ronald-reagan/quotes/above-all-we-must-realize-that-no-arsenal-or-no-weapon.

93 Megan M. Thompson and Rakesh Jetly, "Battlefield Ethics Training: Integrating Ethical Scenarios in High-Intensity Military Field Exercises," *European Journal of Psychotraumatology* 5 (August 2014): 3–5, https://doi.org/10.3402/ejpt.v5.23668; Vallor, "The Future of Military Virtue."

94 Sparrow, "Robots and Respect"; Vallor, "The Future of Military Virtue."

95 Vallor, "The Future of Military Virtue."

96 Cappuccio et al., "A Taste of Armageddon"; Vallor, "The Future of Military Virtue."

97 Lisanne Bainbridge, "Ironies of Automation," *Automatica* 19, no. 6 (1983): 775–79, https://doi.org/10.1016/0005-1098(83)90046-8; Hagendorff, "Mapping the Ethics"; Nalin and Tripodi, "Future Warfare"; Taddeo et al., "Ethical Principles."

98 Mathias Anneken et al., "Ethical Considerations for the Military Use of Artificial Intelligence in Visual Reconnaissance," *arXiv* preprint, 2025, https://arxiv.org/abs/2502.03376; Kearney, "Preserving Human Moral Agency"; Renic, "Tragic Reflection."

[99] Ministry of Foreign Affairs of the People's Republic of China, *China's Position Paper on Regulating Military Application of Artificial Intelligence (AI)*, submitted to the Sixth Review Conference of the High Contracting Parties to the Convention on Certain Conventional Weapons (CCW), UN Doc. CCW/CONF.VI.WP.2, December 13, 2021, https://docs-library.unoda.org/Convention_on_Certain_Conventional_Weapons_-SixthReview_Conference_(2021)/CCW-CONF.VI-WP.2.pdf; Simpson et al., "Militarizing AI: How to Catch the Digital Dragon?"

[100] Khachatryan, "Military AI Challenges"; Kearney, "Preserving Human Moral Agency"; Taddeo et al., "Ethical Principles"; Gregory C. Allen and Isaac Goldston, "The Biden Administration's National Security Memorandum on AI Explained," *Center for Strategic and International Studies*, October 25, 2024, https://www.csis.org/analysis/biden-administrations-national-security-memorandum-ai-explained.

[101] Kearney, "Preserving Human Moral Agency"; Khachatryan, "Military AI Challenges."

[102] De Sio and Van den Hoven, "Meaningful Human Control over Autonomous Systems"; Taddeo et al., "Ethical Principles."

[103] Guo, "The Ethical Legitimacy."

[104] Human Rights Watch and International Human Rights Clinic (Harvard Law School), *Losing Humanity: The Case Against Killer Robots* (New York: Human Rights Watch, 2012), https://www.hrw.org/report/2012/11/19/losing-humanity/case-against-killer-robots; Kelley M. Sayler, "International Discussions Concerning Lethal Autonomous Weapon Systems," *In Focus* IF11294 (Washington, DC: Congressional Research Service, April 19, 2021), https://apps.dtic.mil/sti/pdfs/AD1171922.pdf.