ARTIFICIAL INTELLIGENCE: IMPLICATIONS ON MILITARY DECISION MAKING

Florin OGÎGĂU-NEAMȚIU[1]

The Regional Department of Defense Resources Management Studies (DRESMARA) / "Carol I" National Defense University, Brasov, Romania

Artificial intelligence is fundamentally transforming military decision-making processes across tactical, operational, and strategic levels. This paper examines the multiple implications of AI on military command and control systems, with particular focus on how AI performs across three distinct decision environments: certainty, risk, and uncertainty. Through analysis of military decision theory, AI architectures, and operational applications, this research try to demonstrate that while AI offers unprecedented capabilities in conditions of certainty and calculable risk, it faces significant limitations in the uncertain environments.

Key words: *Artificial intelligence, military environments, risk, war domains, autonomous systems*

1. INTRODUCTION

The rapid advancement of artificial intelligence technologies revealed a new era of military capabilities that extend far beyond traditional weapon systems. From autonomous systems that identify and track targets to predictive analytics that forecast adversary actions, AI systems are increasingly embedded in decisions that determine the course of actions and the allocation of resources. This transformation raises fundamental questions about the nature of military decision-making itself, the role of human commanders in an age of intelligent machines, and the ethical implications of delegating lethal force decisions to automated systems.

Military decision-making has traditionally been understood as a complex cognitive process occurring in what Clausewitz termed the "fog of war" — an environment characterized by incomplete information, time pressure, adversarial deception, and life-threatening stakes. Commanders must make choices based on fragmentary intelligence,

[1] ORCID: N/A, e-mail: florinbvmail@gmail.com

anticipate adversary actions, coordinate complex operations across multiple domains, and accept moral responsibility for outcomes including casualties and destruction. The introduction of AI into this domain represents not merely an enhancement of existing capabilities but an essential shift in how military decisions are formulated, executed, and evaluated.

This paper employs conceptual analysis as its primary methodological approach, drawing on secondary sources including military doctrine, academic literature and published case studies to examine the implications of AI on military decision-making. Rather than generating new empirical data, the research synthesizes existing theoretical frameworks — including decision theory, AI architecture studies, and legal and ethical scholarship — in order to construct an integrated analytical perspective. This approach is particularly suited to an emerging domain where empirical data from live operational environments remains classified or limited, and where conceptual clarity is a prerequisite for sound policy and practice. The limitations of this approach — most notably the reliance on open-source and publicly available materials — simultaneously suggest productive avenues for future research, including empirical studies of AI-assisted decision-making in live military exercises, comparative analysis of national AI defence strategies, and longitudinal assessment of autonomous system performance under operational conditions.

2. MILITARY DECISION-MAKING THEORY

2.1 Classical Military Decision-Making Models

Military decision-making has evolved through centuries of warfare, producing distinct theoretical frameworks that have created the modern command and control systems. Understanding these classical models provides essential context for evaluating how AI impacts military decision processes.

Clausewitz theory emphasizes the fundamental uncertainty and unpredictability of warfare. Carl von Clausewitz's concept of the "fog of war" [1] describes the inherent information deficit commanders face—incomplete intelligence, unreliable communications, and the inability to fully comprehend the battlefield situation. His notion of "friction" captures the countless factors that cause military operations to deviate from plans, from equipment failures to human error or enemy action. Clausewitz argued that while scientific principles apply to certain military problems, warfare ultimately requires genius, intuition, and the ability to instantly grasp the essence of a situation. This emphasis on irreducible uncertainty and the necessity of human judgment stands in tension with AI systems which are based on pattern recognition and statistical prediction.

Boyd's OODA Loop [2] represents a more recent framework that has profoundly influenced modern military thinking. Colonel John Boyd proposed that decision-making in conflict follows a continuous cycle of Observe, Orient, Decide, and Act. The entity that can complete this cycle faster than their adversary gains a decisive advantage, operating inside the opponent's decision loop and creating situations the enemy cannot comprehend or counter in time. The Orient phase, where information is synthesized with experience, culture, and theoretical models to create understanding, represents the most complex and crucial stage. Boyd's framework suggests that speed and adaptability matter more than perfect information or optimal calculations, raising questions about whether AI acceleration of the OODA loop compensates for potential degradation of the orientation process.

Military Decision-Making Process (MDMP) [3] represents the institutionalized approach used by military staffs to analyze problems and develop operational plans. The seven-step process includes receipt of mission, mission analysis, course of action development, course of action analysis (wargaming), course of action comparison, course of action approval, and orders production. This deliberate, analytical approach emphasizes systematic consideration of factors, constraints, and alternatives. MDMP assumes sufficient time for thorough analysis and operates primarily in conditions of risk rather than pure uncertainty, as commanders develop estimates of adversary capabilities and intentions. The structured methodology facilitates staff coordination and commander decision-making but can be time-consuming and may not accommodate rapid adaptation to changing circumstances.

Rapid Decision-Making (RDM) and Intuitive Decision-Making [4] provide alternatives to deliberate analysis in time-constrained situations. Recognition-primed decision-making [5], identified by cognitive psychologist Gary Klein through study of experienced commanders, describes how experts rapidly assess situations by pattern matching to prior experiences, generating a single course of action that they mentally simulate rather than comparing multiple alternatives. This intuitive approach allows rapid decisions based on accumulated expertise but depends critically on relevant experience and can fail when situations differ in crucial but non-obvious ways from familiar patterns.

2.2 Decision Environments: Certainty, Risk, and Uncertainty

A critical framework for understanding AI's role in military decision-making involves distinguishing between three fundamentally different decision environments: certainty, risk, and uncertainty. These categories, rooted in classical decision theory and applied particularly to military contexts, define the nature of information available to decision-makers and the applicability of different analytical approaches.

Decisions under certainty occur when the outcomes of all alternatives are known with complete accuracy. In military operations, such conditions are rare but do exist in constrained scenarios such as ballistic trajectory calculations where physical laws determine outcomes precisely, known terrain navigation where geographic information is complete and accurate, established logistics calculations with fixed transportation networks and resources, and mechanical system operations with deterministic behaviour. In these environments, optimal decisions can be computed algorithmically without ambiguity.

Decisions under risk involve situations where multiple outcomes are possible, but their probabilities can be estimated with reasonable accuracy based on historical data, statistical models, or established patterns. Military decisions under risk include equipment failure predictions based on maintenance histories, weather effects on operations estimated from meteorological models, detection probabilities for different sensor configurations, and casualty estimates for various tactical approaches. Risk-based decisions allow expected value to be calculated, although the exact final outcome remains uncertain.

Decisions under uncertainty characterize situations where neither the full range of possible outcomes nor their probabilities are reliably known. This condition dominates military operations due to adversarial intelligence where opponents actively conceal intentions and capabilities, adaptive adversary behaviour where enemies learn and modify tactics, novel situations without historical precedents, and deception operations designed to mislead decision-makers. The "radical uncertainty" of warfare emerges from the creative, intelligent opposition inherent in armed conflict. Unlike natural phenomena that follow discoverable laws, human adversaries deliberately violate expectations and exploit predictable patterns in their opponents' behavior.

This distinction between certainty, risk, and uncertainty provides a crucial lens for evaluating AI capabilities in military applications, and underpins the analytical structure of the sections that follow [6].

3. AI BASED DECISION MAKING

3.1 Systems in Conditions of Certainty

In certainty-domain operations — where outcomes follow deterministically from known inputs — AI systems deliver their strongest and most reliable performance, as the following examples illustrate.

Fire control and ballistic computation exemplifies AI application under certainty. Modern artillery, naval guns, and precision-guided munitions employ AI-enhanced targeting systems that calculate trajectories accounting for numerous variables including projectile ballistics, atmospheric density and wind patterns, earth rotation (Coriolis effect), temperature effects on propellant, barrel wear characteristics, and target motion predictions. The M982 Excalibur guided artillery projectile achieves error probable of less than 2 meters [7] at maximum range through GPS guidance and trajectory optimization algorithms. The physics governing these calculations is deterministic—given accurate input measurements, the optimal firing solution can be computed with precision extremely difficult to perform by manual calculation. AI systems process sensor data, environmental measurements, and ballistic models in milliseconds, enabling real time responsive fires.

Navigation and route planning in known terrain with functioning positioning systems represents another certainty domain where AI excels. Military logistics and tactical movement require optimal route selection considering terrain trafficability, obstacle locations, enemy threat zones, fuel consumption, and time constraints. Graph algorithms like Dijkstra's shortest path search efficiently optimal routes through complex terrain networks. The Warfighter Machine Interface (WMI) [8] program integrates AI route planning that considers real-time intelligence, known minefields, bridge weight limits, and force positioning to generate movement plans for armored formations. In urban terrain, building layouts and street networks permit algorithmic path planning for infantry movements and unmanned ground vehicle navigation.

Communications network management benefits from AI optimization when network topology and performance characteristics are known. Military tactical networks face challenges of limited bandwidth, intermittent connectivity, and electromagnetic interference.

AI systems can dynamically route traffic, allocate bandwidth, implement Quality of Service priorities, and configure network parameters to maintain connectivity. The WIN-T (Warfighter Information Network-Tactical)[9] system employs automated network management that reroutes traffic around degraded links and optimizes configurations without manual intervention. Given complete knowledge of network state, these optimizations follow from established algorithms.

Maintenance and logistics optimization for equipment with comprehensive sensor monitoring enables predictive maintenance and inventory optimization. The F-35 fighter's Autonomic Logistics Information System (ALIS) [10] collects extensive performance data from aircraft systems, analyses component health, predicts failures, and automatically orders replacement parts. When mechanical failure modes are well-characterized and sensor coverage is adequate, AI can reliably detect degradation patterns and schedule maintenance optimally. Supply chain optimization algorithms minimize inventory costs while ensuring availability, computing optimal stock levels and distribution given deterministic demand forecasts and transportation networks.

Fig. 1 Vehicle Commander (VC) Warfighter-Machine Interface (WMI), consisting of a 180o field-of-view banner across the top, a 60o field-of-view window on the left hand side, and an overhead map on the right hand side. www.researchgate.net

Across all these applications, the shared condition is deterministic input-output relationships that allow AI to outperform humans in speed, precision, and endurance.

3.2. Systems in Conditions of Risk

Where outcomes are probabilistic but statistically estimable — the risk domain — AI systems trained on historical data provide substantial planning and operational value [11].

Intelligence analysis and pattern recognition represents a major application of AI under risk. Intelligence analysts must process vast quantities of data from signals intelligence, imagery, human intelligence, and open sources to identify threats, assess enemy capabilities, and predict adversary actions. AI systems excel at certain aspects of this challenge, particularly pattern recognition in structured data.

Project Maven, initiated by the U.S. Department of Defense in 2017, applies computer vision to analyze full-motion video from drones and other ISR platforms [10]. The system identifies objects of interest including vehicles, buildings, and personnel with accuracy exceeding 90% under good conditions. By automating the laborious task of scanning hundreds of hours of video, AI frees analysts to focus on interpretation and assessment. The system operates under risk rather than certainty because detection depends on probabilistic factors including image resolution, lighting conditions, object orientation, and environmental clutter. Training on large datasets of labelled imagery enables the system to learn statistical patterns distinguishing targets from backgrounds, though performance degrades when conditions differ significantly from training data.

Signals intelligence benefits from machine learning for communications intercept processing, speaker identification, and language translation. Neural machine translation [12] achieves near-human accuracy for high-resource language pairs, enabling rapid processing of foreign communications. Pattern recognition algorithms can identify communications networks, correlate signals to platforms, and detect anomalies indicating new capabilities. These applications work well when processing languages and protocols represented in training data, with performance degrading for rare languages, encrypted content, or novel communications systems.

Threat assessment and predictive analysis employ probabilistic models to estimate adversary capabilities and forecast actions. Bayesian networks can model relationships between observable indicators and underlying threats, updating probability estimates as new intelligence arrives. For example, observing increased fuel deliveries, vehicle movements, and troop concentrations might raise probability estimates that an adversary is preparing an offensive. Time series analysis of historical attack patterns can identify temporal correlations, geographic preferences, and seasonal variations useful for predicting future attacks. AI systems have been employed to forecast IED placement [13], predict areas of instability, and assess terrorist threats based on historical patterns.

However, these predictive systems face significant limitations in military contexts. Adversaries deliberately violate patterns when they become apparent. A predictable enemy is a defeated enemy, so capable adversaries study their opponents' intelligence collection and analysis methods, adapting to evade detection and exploit biases. The baseline assumption of statistical learning—that future data will resemble training data—is systematically violated by intelligent adversaries. This creates fundamental questions about the reliability of AI threat predictions in contested environments.

Operational planning and course of action analysis can leverage AI to evaluate proposed plans under uncertainty. Combat simulations use stochastic models to estimate probable outcomes of operations, running thousands of Monte Carlo iterations to generate probability distributions of casualties, mission success, duration, and logistics consumption. The Synthetic Theatre Operations Research Model (STORM) [14] and similar tools employ AI to evaluate alternative force structures, deployment patterns, and engagement tactics. These models account for weapon accuracy, detection probability, terrain effects, and unit behavior to estimate engagement outcomes.

Reinforcement learning has been applied to tactical decision-making in wargaming contexts, with AI agents learning effective strategies through simulated experience. “AlphaDogfight” [15] an AI system developed by Heron Systems, defeated human F-16 pilots in simulated air combat through reinforcement learning that discovered novel tactics. In simulation, the AI learned to exploit aircraft capabilities and sensor characteristics in ways human pilots had not conceived.

However, the transferability of simulation-learned strategies to actual combat remains uncertain [15]. Simulations necessarily simplify reality, and AI systems optimized for simulation environments may fail when faced with the complexity of actual operations. An AI that learns to win simulated air combat might employ tactics that violate physics, depend on perfect sensor information which are unavailable in reality, or ignore factors not modelled in simulation like pilot stress, equipment malfunctions, and rules of engagement. The risk of "overfitting to simulation" creates questions about how AI-developed tactics translate to actual operations.

Logistics and supply forecasting benefit from probabilistic demand prediction and inventory optimization. AI systems can forecast ammunition consumption, fuel usage, spare parts requirements, and medical supply needs based on historical operations, planned activity levels, environmental conditions, and equipment status. Machine learning models trained on logistics data from Iraq, Afghanistan and Ukraine can predict consumption patterns for future operations. Optimization algorithms can generate supply distribution plans that minimize transportation assets while maintaining adequate supply levels, accounting for probabilistic demand and delivery reliability.

Weather forecasting for military operations employs numerical weather prediction models that are fundamentally probabilistic. Ensemble forecasting generates multiple predictions with varying initial conditions to assess uncertainty and provide probability distributions for temperature, precipitation, wind, and visibility. Military aviation, airborne operations, and amphibious landings depend critically on weather, making accurate probabilistic forecasts valuable for planning. AI-enhanced weather prediction improves forecast accuracy and extends useful forecast horizons.

The effectiveness of AI systems operating under risk depends critically on whether the operational environment resembles the training data and whether underlying probability distributions remain stable.

3.3. AI Systems in Conditions of Uncertainty

In conditions of uncertainty — where neither outcomes nor probabilities are reliably knowable — AI systems face their most significant limitations, for reasons this section examines.

Adversarial intelligence and deception create systematic uncertainty that undermines AI reliability. Deception can be physical (camouflage, decoys, dummy equipment), tactical (feints, demonstrations), or informational (false communications, planted intelligence) while a successful deception creates situations where decision-makers' information is not merely incomplete but actively misleading. Military deception seeks to cause adversaries to act in ways prejudicial to their interests by manipulating their perception of reality.

AI systems trained to recognize patterns are vulnerable to adversarial exploitation when opponents understand those patterns. If an adversary knows that AI systems track vehicle movements to predict offensive operations, they can manipulate vehicle movements to create false indicators while concealing actual preparations. If facial recognition systems rely on certain features, adversaries can develop countermeasures that obscure those features or trigger false matches. The cat-and-mouse dynamic of military competition means that any consistent AI behavior creates exploitable vulnerabilities.

Adversarial machine learning research has demonstrated that neural networks can be fooled by carefully crafted inputs - "adversarial examples" - that appear benign to humans but cause misclassification [16]. Small, imperceptible perturbations to images can cause state-of-the-art image classifiers to confidently misidentify objects.

While these attacks have been demonstrated primarily in laboratory settings, the underlying vulnerability reflects a fundamental limitation: AI systems learn surface statistical patterns rather than developing robust conceptual understanding. An adversary who understands the AI's decision boundary can craft inputs that exploit it.

Novel tactics and adaptation present uncertainty that defeats pattern-based learning. History demonstrates that military innovation often stems from creative employment of existing capabilities in unexpected ways. Blitzkrieg tactics that combined armor, infantry, and close air support in rapid maneuver warfare surprised adversaries prepared for WWI-style static defense. The modern employment of drone warfighting in the Ukraine conflict is evolving constantly to defeat countermeasures, with adversaries adapting triggering mechanisms, rapid research and development cycles, camouflage techniques, and explosive configurations in response to enemy TTPs. Cyber operations employ constantly novel techniques to penetrate defenses and achieve effects. AI systems depend on recognizing patterns seen during training. When adversaries employ genuinely novel tactics without historical precedent, AI systems lack the conceptual understanding necessary to reason about the new situation. A human commander can recognize that "the enemy is doing something unprecedented" and adapt through reasoning from principles. An AI system trained to classify tactics into known categories may force-fit novel behavior into inappropriate categories or simply fail to recognize the significance of observations that don't match learned patterns. The problem is not merely that AI lacks data on novel tactics—it lacks the conceptual framework to understand tactics as purposeful adaptations of means to ends under constraints.

Strategic and political uncertainty surrounding military operations resist quantification and prediction. Military actions occur within political contexts involving competing interests, alliance dynamics, domestic politics, international law, and human psychology. Strategic questions like "Will our allies maintain their commitment if casualties mount?" or "How will the adversary's population respond to bombing?" or "What political objectives actually motivate this conflict?" involve human choices and social dynamics that defy reliable prediction.

Game theory provides some analytical tools for reasoning about strategic interaction, but its application to real conflicts faces severe limitations. Classical game theory assumes rational actors with common knowledge of payoffs and strategies. Real strategic actors have private information, misperceive their situations, act on ideological commitments that are difficult to understand, and make decisions through complex organizational processes rather than as unified actors. AI systems that model strategic interaction through game-theoretic frameworks or learn strategies from historical cases may produce misleading predictions when applied to specific conflicts involving particular leaders, nations, and circumstances.

Emergent behavior in complex operations creates unpredictability even regarding friendly forces. Large military organizations involve thousands or millions of personnel making decisions at multiple echelons in dynamic circumstances. The interactions between tactical actions, logistics constraints, information flow, and human factors produce emergent outcomes that surprise even those within the organization. Morale effects, leadership influence, and small unit initiative can prove decisive but resist prediction from organizational-level data.

AI systems that attempt to predict operational outcomes by modelling component behavior face the complexity problem: comprehensive models become too computationally expensive while simplified models omit crucial factors. The "fog of war" includes uncertainty about one's own forces, not only on enemy sitation. Commanders may not know whether subordinate units will aggressively pursue objectives or proceed cautiously, whether logistics will deliver supplies on schedule, or whether equipment will function reliably. These uncertainties compound, creating situation-dependent outcomes that defy reliable prediction.

4. OPERATIONAL APPLICATIONS ACROSS MILITARY DOMAINS

4.1 Intelligence, Surveillance, and Reconnaissance (ISR)

Intelligence gathering and analysis represent perhaps the most widespread application of AI in military operations, spanning the certainty-risk-uncertainty spectrum with varying degrees of effectiveness across different mission types [17].

Full-motion video analysis from drone feeds generates massive data volumes — a single Predator drone produces the equivalent of a feature-length movie worth of video every 20 minutes of flight. Human analysts cannot possibly review all collected footage, creating what intelligence professionals call the “data deluge” problem. Project Maven demonstrated AI’s capability to address this by automating object detection and tracking in aerial video; convolutional neural networks trained on labelled examples identify vehicles, buildings, and persons with accuracy above 90% under favourable conditions. This application operates primarily in the risk domain, as detection probability depends on measurable factors like image resolution, target size, and environmental conditions.

However, limitations emerge when adversaries employ countermeasures. Camouflage, concealment, and deception can dramatically degrade detection performance. The transition from risk to uncertainty occurs when adversaries understand and exploit AI detection patterns — the cat-and-mouse dynamic of military competition means that static AI pattern recognition faces adaptive adversaries who deliberately violate learned patterns. During the Cold War, the Soviet Union constructed elaborate maskirovka deception operations including fake missile sites and dummy aircraft specifically to deceive reconnaissance systems; modern adversaries employ similar techniques.

4.2 Targeting and Fire Control Systems

Targeting decisions — determining what to strike, when, and with what means — represent one of the highest-stakes applications of military AI, directly involving lethal force with implications for both operational effectiveness and compliance with international humanitarian law.

Target recognition systems can achieve high reliability under certainty and risk conditions. Stationary military installations and large equipment can be identified reliably when image quality is adequate. However, moving to human target identification dramatically increases uncertainty and ethical stakes. Computer vision algorithms can distinguish men from women and children based on size and clothing with reasonable accuracy in controlled conditions, but reliably determining combatant status — the fundamental distinction required by international humanitarian law — exceeds current AI capability. An AI system might identify that an individual is carrying a rifle but cannot determine whether that person is a combatant, civilian hunter, or civilian defending their home [18].

Autonomous weapons systems that can select and engage targets without human control represent the most controversial application. Existing systems range from defensive systems that automatically engage incoming threats (Phalanx CIWS, Iron Dome) to loitering munitions that autonomously search areas for targets matching programmed criteria (IAI Harop, ZALA Lancet). Defensive systems like Phalanx succeed within narrow mission parameters — defending against incoming projectiles involves identifying fast-moving radar returns on intercept trajectories, a relatively deterministic pattern recognition problem. Offensive autonomous weapons in complex environments face fundamental challenges: proportionality assessments that balance military advantage against civilian harm require contextual understanding and ethical judgment that current AI systems lack.

4.3 Logistics and Resource Management

Military logistics involves enormous computational complexity that creates strong opportunities for AI optimization, particularly in certainty and risk domains where outcomes can be predicted reliably.

Predictive maintenance employs AI to forecast component failures before they occur. The F-35's ALIS system exemplifies this approach, collecting performance data from aircraft systems to analyse component health and predict failures. When failure modes are well-characterized and sensor coverage is adequate, this operates reliably in the risk domain. Uncertainty enters with novel failure modes, inadequate training data on new equipment, and combat damage effects — a component may fail in ways not represented in peacetime data, and combat-damaged systems may exhibit erratic behaviour that does not match learned failure patterns.

Supply chain optimization performs well in certainty and risk domains where demand is predictable and transportation networks are secure. However, uncertainty dominates in actual combat operations: an engagement may consume ammunition far faster than predicted, enemy action may sever supply routes, and operational changes may invalidate forecasts entirely. Military logistics requires resilience and redundancy precisely because uncertainty dominates — what matters is maintaining support despite the unexpected, not achieving optimal efficiency under predicted conditions.

4.4 Autonomous Weapons and Lethal Decision-Making

The prospect of weapons systems that select and engage targets without human intervention raises fundamental questions about human control over violence. Drone swarms using decentralized AI coordination represent emerging capabilities with uncertain operational implications. Multiple autonomous drones can coordinate through machine learning algorithms to search areas, overwhelm air defences, or conduct synchronized strikes. China's national defence strategy explicitly prioritizes "intelligentized warfare" involving AI-enabled swarm systems.

The strategic implications extend beyond individual systems. An arms race in autonomous weapons creates risks including lowered thresholds for use of force when autonomous systems enable strikes without risk to personnel, attribution challenges when autonomous weapons create plausible deniability, escalation risks if autonomous systems act faster than human decision-making can assess situations, and potential proliferation to non-state actors. Autonomous weapons represent a transition from the certainty and risk domains where AI performs well into the uncertainty domain where human judgment remains essential — and it is precisely in that domain that the most consequential targeting decisions arise.

5. HUMAN-AI TEAMING IN MILITARY OPERATIONS

Effective integration of AI into military decision-making requires frameworks that leverage AI capabilities while maintaining human judgment. Human-AI teaming models attempt to optimize the complementary strengths of humans and machines across different decision environments.

Manned-Unmanned Teaming (MUM-T)[19] in aviation exemplifies operational human- AI collaboration.

Fig. 2 Apache & Grey Eagle drone

Apache helicopter crews can control unmanned aerial vehicles (UAVs) through the Improved Gray Eagle [20] interoperability system, using drones for reconnaissance while remaining at standoff distances. The human crew tasks the UAV, interprets its sensor feeds, and makes tactical decisions while the autonomous system executes flight control, navigation, and sensor management. This division of labor places certainty and risk-domain tasks (flight control, route following) with AI while preserving human authority over uncertain tactical decisions (target identification, engagement authorization).

The success of MUM-T depends on effective interfaces that provide situational awareness without overwhelming crews, intuitive tasking methods that allow rapid mission updates, and reliable autonomous behavior that maintains crew trust. When autonomy fails unpredictably or interfaces confuse rather than clarify, the human-machine team performs worse than either component alone [21, 22]. Interface design represents a critical challenge, as poorly designed systems create cognitive burdens that offset AI capabilities.

Centaur teaming [23] represents a more integrated model where human and AI capabilities combine throughout the decision process rather than dividing tasks sequentially. The term derives from chess, where "centaur" teams of humans partnered with AI routinely defeat both unaided humans and unaided AI. The human contributes strategic understanding and novel insights while AI provides calculation depth and tactical precision. Applied to military operations, centaur teaming might involve AI rapidly perform data analysis, generate and evaluate courses of action while humans provide contextual judgment, ethical oversight, and strategic coherence.

However, effective centaur teaming requires that humans can critically evaluate AI recommendations—neither blindly accepting nor arbitrarily overriding them. This demands substantial training, well-calibrated AI confidence estimates, and organizational cultures that value human judgment even when it conflicts with algorithmic outputs. The risk of automation bias, where humans defer to AI despite better judgment, threatens to undermine centaur teaming when AI appears authoritative but lacks genuine understanding [24].

Algorithmic decision support positions AI as an advisor that generates recommendations, predictions, or options for human decision-makers who retain ultimate authority. Intelligence fusion systems that integrate multi-source data to present synthesized assessments, predictive analytics that forecast adversary actions, and planning tools that generate candidate courses of action all fall into this category. The human decision-maker receives AI-generated information but is expected to exercise independent judgment.

The effectiveness of decision support depends critically on explainability and trust calibration. If AI recommendations are not understood, humans cannot meaningfully evaluate them and must either blindly trust or ignore the system. If explanations misrepresent how recommendations were actually generated users may develop false confidence. Trust calibration requires that operators understand AI capabilities and limitations, recognizing when recommendations are likely reliable versus when skepticism is to be shown.

Appropriate reliance represents the key challenge in human-AI teaming: fostering appropriate trust rather than over-reliance or under-reliance. Over-reliance leads to automation bias where humans fail to catch AI errors. Under-reliance wastes AI capabilities when valid recommendations are ignored. Appropriate reliance requires operators who understand when AI is likely to perform well (in certainty and risk domains similar to training) versus when it will struggle (in uncertain, novel, or adversarial contexts).

The most effective human-AI teams allocate responsibilities according to comparative advantage across decision environments: AI handles certainty-domain tasks requiring precise calculation at scale, AI supports risk-domain planning with probabilistic analysis and optimization. Humans provide judgment in uncertainty-domain decisions involving adversarial intelligence, novel situations, and ethical considerations, and maintain accountability for all consequential decisions, particularly those involving lethal force.

6. ETHICAL, LEGAL AND GOVERNANCE CHALLENGES

6.1 Algorithmic Bias

While algorithmic bias has been extensively studied in civilian contexts, its manifestation in military AI systems raises distinct concerns with potentially life-and-death consequences. Military AI systems trained on biased data can perpetuate discriminatory patterns in intelligence assessment, targeting, and threat evaluation [25, 26].

Intelligence bias occurs when AI systems trained on historical intelligence data learn patterns that reflect collection biases, analytical assumptions, or discriminatory targeting practices [27]. If training data over-represents certain demographics, regions, or behaviors as threatening due to past biases, AI systems will perpetuate those patterns. Predictive intelligence systems might systematically flag individuals from certain ethnic or religious groups as higher threats based on patterns in biased training data, effectively automating discrimination.

During counterinsurgency operations, if intelligence databases disproportionately associate certain demographic groups with hostile activity due to collection focus or ethnic profiling by human analysts, machine learning systems trained on that data will reproduce those associations. The algorithmic system provides a fake true objectivity that may obscure underlying biases, making discriminatory patterns harder to identify and challenge. Commanders relying on AI threat assessments may unknowingly perpetuate unjust targeting patterns learned from flawed training data.

Sensor bias emerges from the physical characteristics and training data of detection systems. Computer vision algorithms trained primarily on datasets from western populations may perform worse on non-western faces due to training data imbalances. Facial recognition systems have demonstrated significantly higher error rates for women and people with darker skin tones a problem particularly concerning when such systems inform targeting or force protection decisions. If autonomous systems or operators relying on AI assistance experience higher misidentification rates for certain demographics, the risk of wrongful engagement increases.

Thermal imaging and sensor technologies may perform differently across skin tones, clothing types, or environmental conditions more common in certain regions. If detection algorithms are trained and validated primarily in conditions resembling western operating environments, their performance in different geographic or cultural contexts may degrade in ways that disproportionately affect local populations. This technical bias can translate to operational bias with severe humanitarian consequences.

Cultural and linguistic bias affects natural language processing and social network analysis systems. Translation systems trained primarily on formal text may poorly handle dialects, slang, or culturally-specific expressions. Sentiment analysis algorithms developed on a social media may misinterpret communication styles common in other cultures. Network analysis algorithms might misidentify family or tribal relationships as threatening associations based on connection densities common in some cultures but less so in others.

These biases create operational risks beyond ethical concerns. Biased intelligence leads to poor targeting, wasted resources pursuing false leads, and erosion of local population support when innocent people are wrongly identified as threats.

6.2 International Humanitarian Law and Rules of Engagement

International Humanitarian Law (IHL), embodied in the Geneva Conventions and customary international law, establishes fundamental principles governing armed conflict. The use of AI in military operations raises critical questions about compliance with these principles, particularly regarding distinction, proportionality, and precautions.

The principle of distinction requires that parties to conflict distinguish between combatants and civilians, directing attacks only against military objectives. Combatants can be lawfully targeted, civilians cannot be deliberately attacked, and civilian objects cannot be made the object of attack. This seemingly straightforward principle becomes complex in implementation, particularly in irregular warfare where combatants may not wear uniforms and distinction depends on behavior, location, and context.

Current computer vision can identify humans, classify by apparent age and gender, detect weapons, and recognize uniforms with reasonable accuracy under favorable conditions. However, distinguishing combatants from civilians requires understanding of context, behavior, and status that exceeds pattern recognition. A person carrying a rifle might be a combatant, civilian hunter, civilian defending their home, or civilian playing with a wooden rifle similar toy. International law requires positive identification of military status before engagement—determining that someone falls within a targetable category, not merely that they might.

The contextual nature of distinction creates fundamental challenges for algorithmic implementation. A person's status may change over time—civilians who directly participate in hostilities become targetable while participating but regain protected status when participation ceases. Determining whether observed behavior constitutes "direct participation in hostilities" requires judgment about purpose and likely effects that resist algorithmic codification. An AI system might detect that someone is placing an object along a road but cannot determine whether it is an IED (direct participation) or a roadside shrine (protected activity).

Proportionality requires that expected incidental civilian harm not be excessive relative to anticipated direct and concrete military advantage. This assessment involves balancing incommensurable values—military advantage versus civilian casualties—through inherently normative judgment. Different reasonable people can reach different proportionality assessments in identical situations based on their weighting of military necessity versus humanitarian concerns.

Can this judgment be algorithmically automated? Proportionality requires estimating expected civilian casualties (uncertain), assessing military advantage (subjective and context-dependent), and comparing these quantities that resist mathematical comparison. Is destroying a command post worth ten civilian casualties? Twenty? The answer depends on the command post's importance, availability of alternatives, and one's valuing of civilian life against military objectives. These are moral judgments that reflect values, not calculations that have objectively correct answers.

Attempts to quantify proportionality through "casualty value functions" or similar formalisms risk simplifying what should remain contested moral terrain. Reducing proportionality to an algorithm implies that there exists a mathematically optimal exchange rate between civilian deaths and military advantage—a proposition fundamentally at odds with IHL's humanitarian foundations. The requirement for human judgment in proportionality assessment reflects recognition that these are moral choices requiring human moral agency, not technical problems admitting computational solutions.

Precautions in attack require that parties take feasible precautions to minimize civilian harm, including verifying targets, choosing means and methods that avoid or minimize harm, providing warnings when feasible, and canceling attacks when civilian harm would be excessive. These obligations assume human decision-makers who can exercise judgment about verification sufficiency, weigh alternative methods, and assess feasibility of precautions.

Autonomous weapons that engage targets without human oversight at the moment of attack cannot satisfy precautionary obligations in meaningful ways. The system cannot verify targets beyond its programmed recognition criteria, cannot dynamically assess whether civilian harm has become excessive warranting attack cancellation, and cannot provide warnings to civilians. Precautions assume adaptive human judgment responsive to evolving situations—capabilities that autonomous systems lack.

Meaningful human control [28, 29, 30, 31] has emerged as a concept attempting to preserve human agency over violence even as AI systems increasingly mediate military force. What constitutes "meaningful" control remains contested, but proposed frameworks typically require that humans understand how systems function, have sufficient information to make informed decisions, have adequate time to deliberate, and exercise genuine choice. These criteria prove difficult to satisfy as autonomy increases, decision timescales compress, and system complexity grows.

6.3 Command Responsibility and Accountability

Military accountability frameworks assume clear chains of command with commanders responsible for their subordinates' actions. This model faces challenges when AI systems exercise decision-making functions: who bears responsibility when AI makes erroneous targeting decisions, fails to distinguish civilians, or malfunctions in ways causing civilian casualties [32]?

Traditional command responsibility under IHL holds commanders criminally responsible for subordinate war crimes if they knew or should have known about them and failed to prevent or punish them. This doctrine developed in contexts where subordinates are human beings whose actions commanders can observe, predict based on character and training, and control through orders and discipline.

Applying command responsibility to AI systems raises novel questions. Can a commander "know or should have known" about AI system limitations or errors in ways comparable to knowledge about human subordinates? If an AI targeting system misidentifies targets due to training data biases, sensor malfunctions, or adversarial exploitation, does the commander who deployed that system bear criminal responsibility? The commander may lack technical expertise to evaluate AI reliability, may have reasonable belief based on validation testing that the system performs adequately, and may be unable to predict specific failure modes of complex machine learning systems.

Yet allowing AI systems to create accountability gaps would be dangerous. If no human can be held responsible when autonomous weapons commit war crimes, the prospect of punishment that deters violations disappears. Some legal scholars argue that commanders must understand AI systems under their control sufficiently to anticipate foreseeable misuse, while others contend that developers who create systems that are deployed in ways causing violations should bear responsibility. These questions remain unresolved, creating legal uncertainty that may chill beneficial AI applications while failing to prevent harmful ones.

Product liability frameworks developed for defective consumer products provide incomplete analogies for military AI. Product liability requires proving that a product was defective and that the defect caused harm. But what constitutes a "defect" in an AI system? If a computer vision system achieves 95% accuracy, well above human performance on average, but makes an erroneous identification in a specific case, is it "defective"? If the system performs as designed but training data limitations cause it to function poorly in certain conditions, is that a design defect, manufacturing defect, or warning failure?

Military AI systems operate in adversarial environments where enemies actively attempt to defeat them. If adversaries develop countermeasures that exploit AI vulnerabilities causing the system to malfunction, does responsibility lie with developers who created exploitable systems, commanders who deployed them in contested environments, or adversaries who exploited them? The intentional introduction of adversarial stimuli to cause AI errors creates scenarios without clear civilian analogues.

Organizational accountability faces challenges when multiple entities contribute to AI system development, integration, and deployment. Defense contractors develop algorithms, military services integrate them into platforms, training ranges validate performance, operational commanders decide employment, and tactical operators interact with systems in combat. If an AI-enabled targeting error causes civilian casualties, determining which organization's decisions contributed most centrally to the harm may be impossible. The diffusion of responsibility across organizations can create situations where everyone bears partial responsibility but no one bears sufficient responsibility to face meaningful accountability.

6.4 Trust and Human-Machine Interface

Effective human-AI collaboration in military operations requires appropriate trust calibration, and well-designed interfaces. When these elements are lacking, human-machine teams perform worse than either humans or machines alone.

Trust calibration represents a critical challenge. Military operations demand that commanders trust their systems and subordinates as hesitation caused by insufficient trust can be tactically disastrous. However, blind faith in unreliable systems is equally dangerous. Appropriate trust requires that operators accurately understand AI capabilities and limitations, recognize when AI recommendations are likely reliable versus when skepticism is warranted, and maintain vigilance to catch errors when they occur.

Research demonstrates that trust in automation is "sticky" [33] as once established through reliable performance, it persists even when conditions change in ways that degrade reliability. Operators accustomed to AI systems performing well in training or benign environments may fail to increase skepticism when entering contested environments where adversarial action undermines AI effectiveness.

Training programs must expose operators to both AI successes and failures across diverse conditions, establishing realistic expectations about performance boundaries. However, training cannot replicate all conditions that may arise in actual operations. Novel adversary tactics, unusual environmental conditions, or system degradation from battle damage may cause AI failures that no training anticipated. Maintaining appropriate vigilance when AI has proven reliable thousands of times requires cognitive discipline that conflicts with natural human tendencies toward complacency.

Deep neural networks that achieve the highest performance on complex pattern recognition tasks are typically the least interpretable as their decisions emerge from millions of parameters in ways that resist simple explanation. Simpler, more interpretable models often sacrifice accuracy for transparency. In military applications where accuracy can be life-or-death, the performance penalty of interpretable models may be unacceptable. This creates genuine dilemmas between explainability and performance rather than engineering problems admitting technical solutions.

Operational tempo constraints limit explanation complexity. Commanders in time-compressed situations cannot review detailed explanations of algorithmic reasoning—they need rapid assessments of confidence and key factors. However, oversimplified explanations may mislead by suggesting understanding where none exists. Stating "the system is 85% confident” about a valid target based on visual appearance, location, and behavior provides actionable information but obscures the fact that the system could lack genuine understanding of what constitutes a valid target.

Detailed explanations of how AI systems make decisions could enable adversaries to develop countermeasures. If enemies understand how targeting algorithms identify military vehicles, they can modify camouflage, employ decoys, or develop electronic countermeasures exploiting algorithmic weaknesses. The tension between providing operators sufficient explanation to make informed decisions and preventing adversary exploitation of system details creates operational security challenges.

Interface design profoundly influences human-AI team effectiveness. Poorly designed interfaces can overwhelm operators with information, obscure critical factors, or encourage automation bias. Effective interfaces must present AI recommendations clearly without encouraging uncritical acceptance, communicate confidence and uncertainty appropriately, direct operator attention to factors most relevant for decisions, support rapid decision-making without sacrificing critical evaluation, and fail gracefully when AI confidence is low or system malfunctions.

The physical and cognitive interface between humans and AI systems will largely determine whether AI integration enhances or degrades military decision-making. Well-designed interfaces that foster appropriate reliance enable effective human-machine teams. Poorly designed interfaces that encourage over-reliance or fail to support critical thinking create vulnerabilities as dangerous as technical AI limitations.

7. CONCLUSION

This paper has examined the implications of artificial intelligence for military decision-making through the analytical lens of three distinct decision environments. In conditions of certainty, where outcomes follow deterministically from known inputs, AI systems deliver reliable, superior performance — as demonstrated by fire control, route planning, and predictive maintenance applications. In conditions of risk, where outcomes are probabilistic but statistically estimable, AI provides substantial value in intelligence analysis, logistics forecasting, and operational planning, though its effectiveness degrades when adversaries adapt and violate the statistical assumptions. In conditions of genuine uncertainty AI faces fundamental limitations that pattern recognition alone cannot overcome. Human judgment, contextual understanding, and moral reasoning remain irreplaceable.

These findings carry direct implications for military doctrine and force development. Doctrine should formalize a decision-authority framework that allocates tasks to AI or humans based on the prevailing decision environment, rather than treating AI as a uniformly applicable capability. Training programmes must build operators who are calibrated consumers of AI outputs — capable of trusting AI recommendations in appropriate contexts while maintaining critical oversight when conditions shift toward uncertainty. Rules of engagement and command responsibility frameworks require revision to address the accountability gaps that arise when AI systems mediate targeting decisions, ensuring that meaningful human control is preserved not merely in formal policy but in operational practice. Procurement and acquisition processes should incorporate explainability and robustness standards proportional to the decision stakes involved, particularly for systems that operate in or near the uncertainty domain.

Several avenues for future research emerge directly from the limitations of this study. Empirical investigation of human-AI teaming in live or simulated military exercises would test whether the theoretical automation bias risks identified here manifest consistently under operational stress and time pressure. Comparative analysis of how major military powers are institutionalizing AI in doctrine — and whether divergent approaches create interoperability challenges or escalation risks within alliances — represents an urgent strategic research priority. The legal dimensions of command responsibility for AI-enabled systems remain undertheorised and would benefit from systematic doctrinal analysis. Finally, technical research into adversarial robustness — specifically how military AI systems perform when intelligent adversaries deliberately probe their decision boundaries — is essential to close the gap between laboratory performance and operational reliability.

Artificial intelligence is fundamentally transforming decision-making across virtually all domains of human activity including military, creating both extraordinary opportunities and serious challenges that society is only beginning to address. This research has examined the multiple implications of AI integration into decision-making processes through theoretical analysis, case studies, and evaluation of cognitive, ethical, and regulatory dimensions.

AI DISCLOSURE

The author acknowledge the use of the following generative AI tools to assist in the preparation of this manuscript: ChatGPT. This tool was used solely for language editing and structural suggestions, under the complete control and responsibility of the authors. All AI-assisted content was critically reviewed and revised by the authors, who accept full responsibility for the accuracy and integrity of the final version.

REFERENCES

[1] Shapiro, M. J. (2005). The fog of war. Security Dialogue, 36(2), 233-246.

[2] Richards, C. (2020). Boyd's OODA loop. Necesse, 5(1), 142-165.

[3] Marr, J. J. (2001). The military decision making process: Making better decisions versus making decisions better (No. USACGSCSAMSAY200020001).

[4] Klein, G. A. (1993). A recognition-primed decision (RPD) model of rapid decision making. Decision making in action: Models and methods, 5(4), 138-147.

[5] Sauter, Vicki L. "Intuitive decision-making." Communications of the ACM 42.6 1999;

[6] Rashid, M., Abbas, Q., Younis, M. S., Baber, J., & Baber, N. (2023). Artificial intelligence in the military: An overview of the capabilities, applications and challenges. International Journal of Intelligent Systems, 2023, Article 4954561. https://doi.org/10.1155/2023/4954561

[7] Brady, M. R., & Goethals, P. (2019). A comparative analysis of contemporary 155 mm artillery projectiles. Journal of Defense Analytics and Logistics, 3(2), 171-192.

[8] Franks, E. (2005, May). Advanced warfighter machine interface. In Cockpit and Future Displays for Defense and Security (Vol. 5801, pp. 12-23). SPIE.

[9] S. R. Ali and R. S. Wexler, "Army Warfighter Network-Tactical (WIN-T) Theory of Operation," MILCOM 2013 - 2013 IEEE

Military Communications Conference, San Diego, CA, USA, 2013, pp. 1453-1461, doi: 10.1109/MILCOM.2013.246.

[10] https://www.lockheedmartin.com/content/dam/lockheed-martin/rms/documents/alis/CS00086-55%20(ALIS%20Product%20Card).pdf

[11] Nadibaidze, A., Bode, I., & Zhang, J. (2024). AI in military decision support systems: A review of developments and debates. Center for War Studies, University of Southern Denmark. https://doi.org/10.21996/epbr7566

[12] Lundberg, S.M., Lee, S.I., A Unified Approach to Interpreting Model Predictions, Advances in Neural Information Processing Systems, 2017.

[13] Lerner, W. D. (2013). Predicting the emplacement of improvised explosive devices: an innovative solution. Capitol College.

[14] Lucas, T. W., Sanchez, S. M., Sanchez, P., McDonald, M., & Upton, S. (2015). Developing Synthetic Theater Operations Research Model (STORM) Analytic Utility, Phase II.

[15] Pope, A. P., Ide, J. S., Mićović, D., Diaz, H., Rosenbluth, D., Ritholtz, L., ... & Javorsek, D. (2021, June). Hierarchical reinforcement learning for air-to-air combat. In 2021 international conference on unmanned aircraft systems (ICUAS) (pp. 275-284). IEEE. [] Doshi-Velez, F., Kim, B., Towards A Rigorous Science of Interpretable Machine Learning, ArXiv preprint, 2017.

[16] Papernot, N., McDaniel, P., Swami, A., & Harang, R. (2016, November). Crafting adversarial input sequences for recurrent neural networks. In MILCOM 2016-2016 IEEE Military Communications Conference (pp. 49-54). IEEE.

[17] Borghoff, U. M., Biundo, S., Schmid, U., & Thielscher, M. (2025). The use of artificial intelligence in military intelligence: An experimental investigation. Frontiers in Human Dynamics, 7, Article 1512697. https://doi.org/10.3389/fhumd.2025.1512697

[18] Viveros Álvarez, I. (2024). The risks and inefficacies of AI systems in military targeting support. ICRC Humanitarian Law & Policy Blog. https://blogs.icrc.org/law-and-policy/2024/05/14/risks-inefficacies-ai-systems-military-targeting-support/

[19] Taylor, G., & Turpin, T. (2015). Army aviation manned-unmanned teaming (MUM-T): past, present, and future. In 18th International -Symposium on Aviation Psychology (p. 560).

[20] Judson, J. (2014). Improved Gray Eagle Shows Greater Endurance With Heavier Payloads. Inside the Army, 26(21), 10-10.

[21] Horowitz, M. C., & Lin-Greenberg, E. (2024). Bending the automation bias curve: A study of human and AI-based decision making in national security contexts. International Studies Quarterly, 68(1), sqad100. https://doi.org/10.1093/isq/sqad100

[22] Jonsson, M. (2023). Trusting machine intelligence: Artificial intelligence and human-autonomy teaming in military operations. Defence Studies, 23(3), 439–458.

https://doi.org/10.1080/14702436.2023.2178154

[23] Johnson, James, The AI Commander: Centaur Teaming, Command, and Ethical Dilemmas (Oxford, 2024; online edn, Oxford Academic, 18 Jan. 2024), https://doi.org/10.1093/oso/9780198892182.001.0001, accessed 13 Nov. 2025.

[24] Abioye, S. O., Busari, H. A., Omole, D. O., John-Mathews, J.-M., & Aremu, B. (2024). AI-driven human-autonomy teaming in tactical operations: Proposed framework, challenges, and future directions. arXiv preprint arXiv:2401.12172.

[25] Pace, T., & Ranes, B. (2024, June). Bias, explainability, transparency, and trust for AI-enabled military systems. In Assurance and Security for AI-enabled Systems (Vol. 13054, pp. 20-29). SPIE.

[26] Bode, I., & Bhila, T. (2024). The problem of algorithmic bias in AI-based military decision support systems. ICRC Humanitarian Law & Policy Blog. https://blogs.icrc.org/law-and-policy/2024/09/03/algorithmic-bias-ai-military-decision-support-systems/

[27] Parasuraman, R., Manzey, D.H., Complacency and Bias in Human Use of Automation: An Attentional Integration, Human Factors, Vol. 52, No. 3, 2010.

[28] Cavalcante Siebert, L., Lupetti, M. L., Aizenberg, E., Beckers, N., Zgonnikov, A., Veluwenkamp, H., ... & Lagendijk, R. L. (2023). Meaningful human control: actionable properties for AI system development. AI and Ethics, 3(1), 241-255.

[29] Greipl, S. (2024). AI in military decision-making: Supporting humans, not replacing them. ICRC Humanitarian Law & Policy Blog. https://blogs.icrc.org/law-and-policy/2024/08/20/ai-military-decision-making-supporting-humans/

[30] Docherty, B., & Kuster, M. (2023). Regulating lethal autonomous weapon systems: Exploring the challenges of explainability and traceability. AI and Ethics, 3(4), 1425–1438. https://doi.org/10.1007/s43681-022-00261-6

[31] ICRC & Geneva Academy of International Humanitarian Law and Human Rights. (2024). Expert consultation report: AI and related technologies in military decision-making on the use of force in armed conflicts. Geneva: ICRC. https://www.icrc.org/en/document/ai-related-technologies-military-decision-making-use-force-armed-conflicts

[32] Nalin, F., & Tripodi, P. (2023). Future warfare and responsibility management in the AI-based military decision-making process. Journal of Advanced Military Studies, 14(1), 58–76. https://doi.org/10.21140/mcuj.20231401003

[33] Liang, G., & Newell, B. (2022). Trusting algorithms: performance, explanations, and sticky preferences. In Proceedings of the annual meeting of the cognitive science Society (Vol. 44, No. 44)

[34] Wilson, N. A. (2020). Understanding the Battle for AI in Warfare through the Practices of Assemblage: A Case Study of Project Maven (Master's thesis)