

ORIGINAL RESEARCH

Open access

Machine Learning in Defense and Military Applications: Opportunities, Challenges, and Ethical Considerations

Lucas Meyer^{1*}, Anna Schmidt², Stefan Braun¹, Laura Keller³

Abstract

Machine learning (ML) has evolved into an indispensable technology within the defense sector, driving the creation of sophisticated mechanisms for identifying threats, supporting decision-making processes, and managing autonomous operations. Nevertheless, the expanding integration of ML into defense frameworks has generated significant ethical anxieties concerning accountability, transparency, and inherent bias. This paper offers an extensive evaluation of ML's influence on the defense industry, detailing both the advantages and risks of implementing these methodologies across diverse domains, including surveillance operations, target pinpointing, and autonomous weapon platforms. Furthermore, we examine the ethical repercussions of deploying ML in military contexts, emphasizing problems tied to privacy preservation, responsibility attribution, and algorithmic bias. Ultimately, we outline strategic recommendations to alleviate these ethical complications, advocating for heightened operational transparency, clearer accountability structures, and the active engagement of diverse stakeholders in the architecture and deployment of ML systems within defense frameworks.

Keywords Machine learning, Artificial intelligence, Defense, Military tactical environments, Industry

*Correspondence:

Lucas Meyer

lucas.meyer@ethz.ch

¹ Department of Defense AI and Machine Learning, Faculty of Engineering, ETH Zurich, Zurich, Switzerland

² Department of Military Applications and Ethical AI, Faculty of Engineering, University of Bern, Bern, Switzerland

³ Department of Defense AI Opportunities and Challenges, Faculty of Engineering, EPFL Lausanne, Lausanne, Switzerland

Introduction

Over the past several decades, the concept of machine learning (ML) has achieved extensive recognition, expanding far past its origins in the academic and scientific communities. ML is frequently promoted as an accessible tool to optimize any framework into which it is incorporated. Shifting away from standard classical programming, ML capitalizes on modern technological breakthroughs—specifically heightened computing power and the accumulation of massive datasets (Big Data)—to streamline the development of algorithms capable of mapping the relationships between inputs and outputs. Consequently, these algorithmic models can be leveraged to compute statistical predictions regarding the probability of future occurrences.

Methodologies and tools grounded in ML, or more broadly within the domain of artificial intelligence (AI), have fundamentally altered numerous industrial sectors by offering substantial benefits, which include:

- **Improved efficiency:** ML models possess the capability to automate routine, labor-intensive operations like data input and analysis, thereby liberating human personnel to dedicate their efforts to more intricate and innovative assignments [1].

- **Increased accuracy:** ML algorithms are capable of interpreting substantial datasets with high speed and precision, uncovering subtle patterns and insights that would remain obscured to human analysts [2]. This capability allows commercial enterprises to formulate better-informed choices, optimize product quality, and minimize operational errors.
- **Cost savings:** Through the dual mechanisms of task automation and error reduction, ML enables businesses to decrease expenditures related to manual labor and lower overall operational waste [3].
- **Predictive maintenance:** By processing data gathered from sensors and historical telemetry, ML can forecast when operational machinery is prone to malfunction, empowering enterprises to execute preventative maintenance routines before catastrophic breakdowns take place [4].
- **Fraud detection:** ML frameworks excel at identifying anomalous data configurations that point toward deceptive behavior, such as credit card manipulation or insurance fraud schemes [5].
- **Improved supply-chain management:** Through the multi-tiered analysis of logistics networks, ML models can pinpoint operational bottlenecks, allowing organizations to decrease inventory overhead and optimize delivery schedules [6].

At present, a multitude of prominent global enterprises, including Google, Netflix, Amazon, Twitter, Facebook, IBM, Apple, Microsoft, and Oracle, are allocating massive capital reserves toward ML innovations to evaluate user demographics and construct next-generation products for the marketplace. As an illustration, Netflix utilizes machine learning frameworks to track and analyze consumer search queries, thereby tailoring the media recommendations displayed on user landing screens to align with individual tastes [7]. Similarly, Amazon's Alexa architecture captures user vocal inputs and transmits this data to the cloud-based Alexa Voice Services, where ML algorithms interpret the underlying verbal directives to return appropriate responses to the host device [8]. These implementations effectively transition technologies originally engineered for civilian environments into tactical or defense settings, as seen in the military adaptation of virtual assistants (such as Apple's Siri or Amazon's Alexa), voice recognition systems, and speech-to-text tools for transcription and translation utilities. On the whole, the marriage of ML and AI yields critical advantages for commercial and industrial landscapes, including the defense sector, particularly when viewed through a security lens. The integration of complementary systems alongside AI and ML within defense architectures was evaluated in [9], highlighting innovations like neuromorphic hardware designed for accelerated processing of high-volume data streams. Furthermore, novel and modernized tactical deployment scenarios leveraging these emerging technologies were conceptualized in [9, 10].

In tandem with these developments, the defense domain shares a structural alignment with Public Protection and Disaster Relief (PPDR) operations and Mission-Critical Services (MCSs), stemming from their collective mandate to maintain public safety and national security. Both defense frameworks and PPDR/MCS networks are deeply embedded in emergency containment and crisis interventions, working collaboratively to minimize the fallout of natural or anthropogenic catastrophes and shield civilian populations from danger. Additionally, military assets and PPDR/MCS agencies share the responsibility of protecting critical infrastructure [11], such as electrical grids, communication links, transit networks, state facilities, and water distribution systems. This defensive posture is vital for ensuring the continuity of core societal functions during active emergencies [12]. In this framework, ML models and broader artificial intelligence architectures can stimulate the technological breakthroughs necessary to foster institutional collaboration, resulting in synchronized responses and heightened resilience against multifaceted threats. Prominent investigations, such as the work by Petrov *et al.* in [13] regarding the enforcement of end-to-end dependability for mission-critical traffic within 5G network software, the study by Spantideas *et al.* in [14] evaluating smart Mission-Critical Services across Beyond 5G networks with an emphasis on feedback loops and predictive overload management, and the research by Skarin *et al.* in [15] targeting edge-based mission-critical management over 5G architectures, illustrate the vanguard of research in this domain. These inquiries laid the foundation for embedding sophisticated ML architectures into the core operational pipelines of the defense industry. It is evident that ML systems are critical to optimizing these workflows, given that their inherent flexibility and computational power allow them to be directly translated to defense environments, markedly amplifying operational parameters. Furthermore, ML applications were explored in [16] regarding a specific utility designated as DIVVA, which is designed to authenticate and validate crisis-related data proliferated across social

networking channels. This diagnostic system deployed an ML methodology centered on a bidirectional LSTM framework, yielding an 84% accuracy rate in data categorizations.

In [17], a comprehensive taxonomy of ML approaches applicable to disaster identification is organized across six primary domains: early-stage damage warnings, structural damage quantification, surveillance and identification, prognostic modeling and forecasting, post-disaster response and asset orchestration, and extended risk mitigation and assessment. According to [17], artificial neural networks represent the most capable contemporary strategy for seismic activity tracking. Conversely, more traditional Support Vector Machines (SVMs) have been successfully deployed in research settings to capture variations across imagery, thereby facilitating automated classification workflows. In a similar vein, an array of mission-critical functions is materializing to support critical operations and infrastructure. Within public cellular networks relying on 5G and nascent 6G architectures, centralized deep reinforcement learning (CDRL) along with federated DRL (FDRL) serve as viable ML options for managing critical workflows [18].

Nevertheless, whenever these technological remedies intersect with the defense industry, they must be rigorously evaluated from ethical and jurisprudential vantage points due to the imperative of human oversight. It is precisely within this context that our current study is situated.

Our Contribution

The widespread integration of machine learning (ML) has fundamentally reshaped a multitude of commercial sectors, and the defense industry serves as a primary example of this operational transition. Even so, the military domain demands a much higher level of scrutiny because of the severity of its operations and the direct consequences that systemic decisions impose on human survival. Consequently, while the existing literature features numerous studies outlining the deployment of artificial intelligence and machine learning within tactical frameworks or military hardware, there is an absence of research dedicated to evaluating the legal and ethical ramifications of these systems. To bridge this gap, the current study expands upon the previous insights established by the authors in [19].

The specific contributions of this paper are organized as follows:

- An interrogation of the ethical and legal complexities introducing ML algorithms and merging them into active defense operations.
- A structural blueprint detailing the necessary parameters of a governance framework designed to oversee ML-defense orchestration from a legal and ethical standpoint.
- An analytical breakdown of the technical hurdles, merits, and liabilities of embedding ML within defense networks, supplemented by an evaluation of an active project that has integrated ML methodologies.

To conclude, the introductory section offers a concise distillation of the accelerating momentum toward utilizing ML assets for military purposes, establishing the core purpose of this paper: to identify and unpack the latent vulnerabilities and systemic obstructions inherent to this transition.

Materials and Methods

The adopted methodology was rooted in the authors' substantial military background and direct engagement with operational ML frameworks. The goal was to identify and analyze the structural challenges of embedding ML within military operations, thereby facilitating a more robust decision-making paradigm for its effective deployment in this domain.

- **Data Collection:** Information was gathered using a hybrid approach consisting of structured technical literature reviews and active project participation. The authors consulted with military colleagues, defense technology developers, and ML practitioners to collect empirical insights and qualitative observations regarding implementation

obstacles. Due to security classifications inherent to military environments, these projects are not described in full detail.

- **Case-Study Selection:** A specific case study evaluating multiple ML deployments within the defense industry was chosen. This case study focused on the SALAs initiative and integrated empirical examples from a broader survey concerning perceptions of lethal autonomous weapon systems.
- **Problem Identification:** The accumulated data underwent systematic categorization to isolate recurring technical difficulties observed across our combined experiences. These challenges were classified into technical, ethical, operational, and strategic vectors to ensure a multidimensional perspective.
- **Comparative Analysis:** A comparative evaluation of the identified challenges was executed to establish a conceptual framework. While these issues could be broadly attributed to artificial intelligence as a whole, the analytical focus remained strictly on their military implications.

Leveraging their shared operational experience, the authors formulated practical recommendations aimed at alleviating these challenges and advancing the successful adoption of ML in defense frameworks. These strategies address legal and ethical dimensions that have hitherto been overlooked in existing academic literature and research.

ML and Defense Sector

The utilization of ML within defense frameworks possesses the capacity to augment security, optimize command decisions, and elevate operational performance. The civilian applications outlined previously can be structurally adapted to military environments. This integration would completely reshape the tactical landscape and drive the adoption of cutting-edge technologies. This operational paradigm shift was detailed in [9], which articulated ten specific use cases involving emerging architectures like 5G and nascent 6G networks. Although ML algorithms were conceptualized decades ago, their practical implementation within real-world systems was previously constrained. Today, 5G and forthcoming 6G communication frameworks introduce optimization techniques that enhance data processing capabilities and expand accessible data volumes, thereby making the deployment of ML across diverse systems feasible. Consequently, it is critical to resolve the dilemmas introduced by the integration of ML within the defense sector.

Virtual assistants capable of processing and executing vocal directives allow for hands-free device management while delivering automated responses to routine inquiries. This capability directly reinforces and safeguards the operational security of soldiers deployed in theater. Similarly, speech-to-text algorithms convert spoken communication into text, facilitating real-time transcriptions of briefs, lectures, and operational debriefs, which subsequently enhances situational communication and improves human safety parameters. Nevertheless, adapting these civilian systems for military deployment introduces profound legal and ethical challenges that demand rigorous analysis. Evaluating the legal and ethical ramifications of AI and ML within the defense sector is critical for several key reasons.

First, the deployment of these technologies in military contexts is governed by national and international legal frameworks that regulate weapon design, production, and field use. For example, the engineering and deployment of autonomous weapon systems create significant friction regarding accountability, operational transparency, and the maintenance of human control. Resolving these questions is imperative to guarantee compliance with international humanitarian law and preserve global stability and peace.

Second, the utilization of AI and ML in defense systems can yield unintended operational outcomes, such as systemic biases, systemic errors, and discriminatory outputs. As an illustration, facial recognition models integrated into defense configurations can replicate and reinforce racial and gender biases. It is therefore vital to construct a legal framework that proactively accounts for the latent risks and downstream harms linked to these technologies.

Third, ethical considerations must serve as a foundational element when implementing AI and ML in defense. These systems prompt fundamental inquiries regarding the moral accountability of individuals and institutions, the sanctity of

human life, and the preservation of human rights. For instance, relying on AI and ML within decision-making pipelines may yield conclusions that conflict with foundational ethical principles such as fairness, equity, and justice.

Despite the clear imperative to construct robust legal structures and address ethical challenges, significant deficiencies persist. A major portion of current legal frameworks remains outdated and fails to fully capture the distinct problems introduced by AI and ML deployment in military environments. Furthermore, ethical parameters are frequently sidelined, and there remains a lack of international consensus regarding the specific ethical doctrines that should steer the development and application of these systems.

Drawing upon the fundamental operational principles of ML and its current commercial utilization by global enterprises, this study evaluates the ethical and legal issues tied to deploying this technology within national defense initiatives, alongside its associated liabilities, structural challenges, and operational benefits. It must be recognized that when analyzing national defense or armed conflict, analysis cannot be confined solely to physical, destructive engagements involving tangible material assets. The scope must encompass cyber-attacks, network breaches, terrorism, public anxiety, societal insecurity, maritime piracy, unregulated migration, space-based threats, and any other action or omission capable of generating a state of national hostility.

A series of distinct advantages and disadvantages regarding the implementation of ML across various defense architectures has been synthesized, as illustrated in **Figure 1**.

The convergence of ML with defense technologies yields several key advantages. First, it facilitates the detailed simulation of potential conflict environments, enabling defense agencies to rigorously test diverse strategies and construct effective counter-measures. Furthermore, machine learning enhances troop training methodologies by generating realistic, dynamic simulation environments that optimize decision-making capabilities and situational awareness under stress. In addition, ML algorithms can process historical data streams to extract critical operational insights, thereby assisting personnel in understanding and responding to dense contextual situations more effectively. By synthesizing massive volumes of data, these algorithms provide operators with augmented, actionable intelligence, supporting command choices and elevating overall mission efficiency. Another major benefit is the capacity for early threat detection and the mitigation of future crises. These algorithms can identify subtle patterns and structural anomalies within incoming data to flag latent dangers, allowing for the execution of proactive measures to neutralize threats before they materialize.

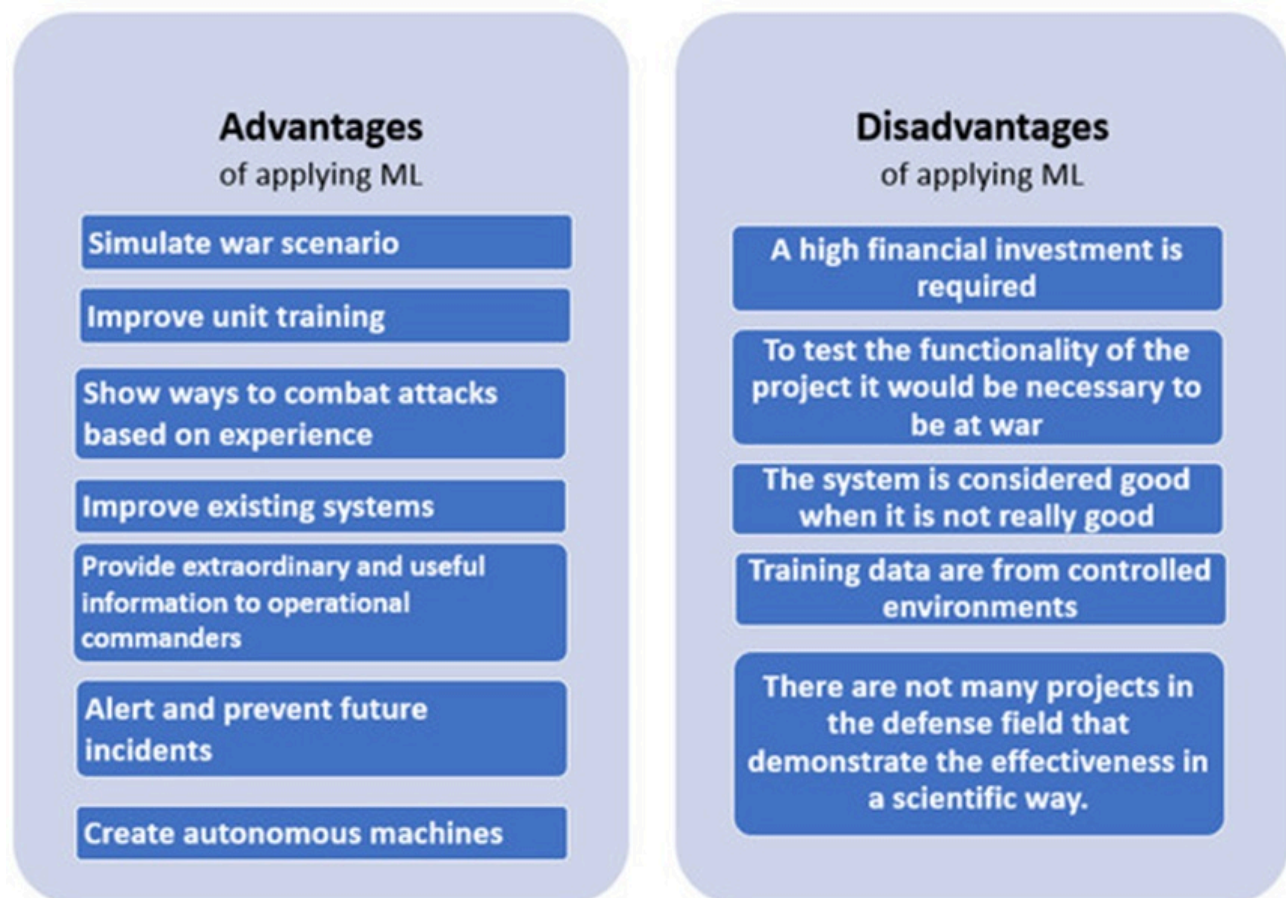


Figure 1. Advantages and disadvantages of applying ML in military environments.

Conversely, significant drawbacks must be addressed. Integrating machine learning into defense architectures demands substantial financial investments, as it requires constructing and maintaining resilient hardware infrastructure, acquiring cutting-edge technology assets, and retraining personnel. Moreover, sourcing relevant, high-fidelity datasets to train these machine learning algorithms remains an operational hurdle, given that training requires extensive and diverse data inputs that precisely mimic real-world tactical environments.

Defense Systems for Integrating ML

Multiple defense frameworks currently incorporate technological architectures driven by ML. The primary systems identified as pioneers in leveraging these methodologies to advance military technologies are compiled below:

- 1. Autonomous systems:** This category encompasses unmanned aerial vehicles (UAVs) and unmanned ground vehicles, which utilize ML algorithms to navigate environments and execute operations independent of direct human steering. For instance, Ref. [20] introduced an autonomous network framework relying on multi-agent reinforcement learning for proactive threat identification—a capability that has grown vital to the contemporary cybersecurity ecosystem due to the accelerating frequency and scope of cyber intrusions.
- 2. Predictive maintenance:** ML processing models are utilized to evaluate data streams from embedded sensors and external telemetry to project when equipment is susceptible to failure, thereby facilitating repair operations before a complete breakdown occurs. The SOPRENE project [21] introduced the application of ML explicitly for predictive maintenance functions.

3. **Cybersecurity:** ML mechanisms can monitor and analyze network data packets to isolate behavioral anomalies and prospective intrusions, facilitating accelerated countermeasure deployment and decreasing the vulnerability to cyber attacks [22].
 4. **Situational awareness:** ML models can synthesize heterogeneous inputs from a wide variety of channels, including tactical sensors, optical cameras, and social media streams, to deliver real-time operational context to field personnel [23]. Furthermore, the automated detection of refugee settlements from satellite imagery using multi-class graph-cut segmentation alongside shadow projection metrics was demonstrated in [24].
 5. **Logistics and supply-chain management:** ML computational models can streamline logistical processes and supply networks by computing correlations across inventory levels, transit timelines, and adjacent variables to maximize operational efficiency and curtail expenditures [25].
 6. **Threat detection:** ML tracking models can be applied to identify potential hazards, such as explosive materials or illicit weaponry, at tactical security checkpoints or during freight cargo evaluations [20].
- On the whole, ML-driven technologies have the capacity to substantially augment the structural capabilities of defense infrastructures and defense personnel, leading to advancements in efficiency, precision, and operational safety.

Legal Framework

Historical patterns demonstrate that as pioneering technologies materialize and achieve widespread public adoption, legal and political systems step in to govern their utilization. These interventions frequently manifest as ethical, health-focused, or societal regulatory frameworks. The evolution of ML methodologies and their capacity to execute automated determinations independent of human oversight is no exception to this trend; however, formal regulatory considerations have only emerged in recent years as these systems transitioned from abstract academic concepts into real-world applications. The European Commission (EC) issued a landmark "White Paper" in Brussels in February 2020 [26], which served as the inaugural foundational document addressing the statutory regulation of AI on a European scale. The core objective of this initiative was to cultivate an ecosystem of excellence capable of supporting the engineering and assimilation of artificial intelligence throughout the EU economic landscape and public administrations. This was intended to be achieved by capitalizing on the advantages of professional and industrial markets alongside the massive quantities of global digital data being generated.

According to an authoritative press release by the EC [27], any future legislative framework governing AI must satisfy seven core requirements:

- Human action and oversight.
- Technical soundness and security.
- Privacy and data management.
- Transparency.
- Diversity, non-discrimination, and fairness.
- Social and environmental well-being.
- Accountability.

Furthermore, profound concerns have been articulated regarding the preservation of fundamental human rights, specifically personal data protection, individual privacy, and non-discrimination. Equal emphasis has been placed on defining the civil and criminal liabilities associated with operations executed by autonomous machinery or platforms operating via machine learning.

Grounded in these regulatory considerations, a member nation of the European Commission initiated a pioneering AI regulatory pilot program designated as a "sandbox" in June 2022 [28]. This sandbox operates as an operational mechanism to execute an initial regulatory prototype, drawing directly upon the shared experiences of legislative

authorities and AI development enterprises to catalog best practices for the deployment of this technology. This pilot initiative is projected to serve as the baseline architecture for the forthcoming European Regulation on AI, with full implementation anticipated within the subsequent two years.

Ethical Framework

The ethical problems introduced by AI architectures and the deployment of ML algorithms have been explored across numerous domains. For example, UNESCO evaluated the ethical challenges of AI within diverse spaces, including the automotive, legal, and artistic sectors in [29]. Nevertheless, the defense sector has historically been excluded from these frameworks, which underscores the necessity of our present analysis. A primary objective within modern policy is satisfying the Sustainable Development Goals [30]; correspondingly, the defense industry and its tactical applications must scrutinize these benchmarks and evaluate their internal operational impacts.

Undeniably, the deployment of combat architectures must strictly align with International Humanitarian Law (IHL), which dictates a body of regulations engineered to restrict the collateral effects of armed conflicts based on humanitarian imperatives (Hague Law). To achieve this compliance, it is vital that human commanders maintain oversight over military actions to exercise a context-appropriate degree of discrimination and caution dictated by the tactical environment. This oversight ensures that the hazards imposed on non-combatants remain proportional to the strategic importance of the military objective. However, in active combat environments where an adversary might field lethal autonomous weapon systems, the semi-automatic configuration of these operational thresholds may conflict with the survival of units deployed on the battlefield and the overall success of the mission.

While military AI can be highly valued for its processing speed, decisiveness, and precision, it simultaneously introduces severe risks due to its capacity to finalize conclusions without human temperance, ethical reflection, or fluid adaptation to the volatile complexities of active warfare. Variables such as distinguishing between active combatants, peaceful populations, civilians, and coalition forces; executing strikes with disproportionate force; systemic vulnerabilities; identity spoofing; and the deployment of inadequately vetted field systems could inadvertently accelerate military escalation and heighten ambient danger rather than yielding a strategic edge. It remains an extraordinary challenge for an automated machine to independently comprehend and calculate the principles of distinction and proportionality. For instance, distinguishing between a terrified civilian and a hostile combatant in a dense urban theater, or calibrated defensive force in response to a physical assault, demands a complex quantitative, qualitative, and ethical critique.

Additional pillars of concern involve accountability structures and the preservation of human dignity. The assignment of responsibility functions as a critical deterrent against reckless actions, ensures adherence to legal standards, and provides moral accountability. Human dignity is equally paramount from an ethical perspective; a human casualty decided exclusively by an autonomous algorithm risks reducing human life to a mere computational asset. The choice to terminate a human life should, at an absolute minimum, necessitate a prior moral evaluation.

A high-profile instance of the ethical friction generated by integrating ML into tactical military settings is Google's decision to terminate its partnership regarding Project Maven with the US Pentagon. This choice proved highly contentious, sparked by over 4,000 corporate employees signing a petition demanding the dissolution of the contract, signaling a profound deficit of trust regarding state utilization of the system [31].

Lastly, it is important to note that in 2015, the Open Roboethics Institute (ORI) orchestrated a global opinion poll that captured perspectives from over 1,000 respondents spanning 54 separate nations. Among various thematic queries, this investigation cataloged global perspectives regarding the deployment of lethal autonomous weapon systems (SALAs) [32]. The empirical conclusions derived from the SALAs survey [32] are aggregated and presented in **Table 1**.

Table 1. Opinion survey on the use of SALAs.

Response Percentage	Reworded Statement
67%	Respondents indicated that every category of SALAs ought to be subject to an international ban.
56%	Participants supported prohibiting both the development and deployment of SALAs.
85%	Respondents agreed that SALAs should not be employed in offensive military operations.
71%	Participants preferred the use of remotely controlled weapon systems as an alternative to SALAs.
60%	Respondents stated that, if subjected to an attack, they would rather face a remotely operated system than a SALA.

Reviewing the data organized within [Table 1](#), it is apparent that from an ethical standpoint, the development and active fielding of SALAs faced substantial opposition from the surveyed population. Specifically, 67% of the participants asserted that every category of SALAs ought to be prohibited by international treaty. Moral and ethical considerations clearly eclipsed military expediency and operational optimization, effectively challenging the historic adage that "all is fair in war."

Example of ML Application in Defense: ATLAS

The Advanced Targeting and Lethality Automated System (ATLAS) initiative [\[33\]](#) was primarily established to integrate artificial intelligence and machine learning functionalities into United States combat tanks. This integration enables these vehicles to detect and engage targets at three times the speed of traditional techniques. Consequently, from an AI and ML perspective, the ATLAS initiative centered on the creation of a learning algorithm optimized to interpret massive volumes of sensor-derived data. The objective was the automated detection and classification of hazards, alongside the automatic calculation of the necessary azimuth and elevation for weapon systems to execute an engagement [\[34\]](#). Furthermore, the system incorporated an ML algorithm designed to directly allocate the most appropriate available weapon to neutralize the identified threat. To accomplish these goals, ATLAS prioritized advancements within the following technical domains:

- **Data collection and algorithm pre-training:** Gathering comprehensive datasets regarding potential categories of military targets to facilitate the initial training of the utilized ML algorithm.
- **Image processing:** Enhancing capabilities related to target detection, classification, recognition, identification, and tracking through the implementation of dedicated ML methodologies.
- **Trigger control:** Developing sophisticated targeting algorithms, automating the engagement sequence, and generating optimal weapon recommendations based on the specific nature of the identified target.
- **Vehicle technical integration:** Adapting the combat vehicle's internal infrastructure to support a high-voltage power architecture (600 Vdc) alongside the necessary sensor arrays and electronic components.
- **Sensors:** Equipping combat tanks with advanced hardware to enable real-time data streaming for the ML algorithm. This includes imaging sensors spanning the visible, NIR (near-infrared), SWIR (short-wave infrared), MWIR (medium-wave infrared), and LWIR (long-wave infrared) spectrums; gyroscopic platforms facilitating continuous 360° rotation for the sensor suite and rangefinder; and LADAR (laser detection and ranging) or LIDAR (light detection and ranging) systems.

Challenges in the Defense Sector

While ML offers substantial promise for military applications, engineering dependable defense products still necessitates a heavy reliance on simulation frameworks, expert knowledge, and data engineering, as noted in reference [35]. The primary difficulties identified during this assessment are outlined in **Figure 2**.



Figure 2. Challenges of applying ML in military environments.

- **Potential for friendly fire:** Incidents of fratricide may occur due to the incorrect identification of friendly assets, confusion between allied and adversarial forces, communication failures regarding asset status, or inadequate contextual awareness during target development. Moreover, automated task execution frequently lacks tactical patience or deliberate pre-action reflection, potentially leading to unverified collateral damage.
- **Adversarial manipulation of ML models:** The widespread adoption of these technologies over time has driven the emergence of techniques designed to exploit or disrupt them (adversarial evasion attacks). This has forced researchers to rigorously evaluate the reliability, privacy, and security of these architectures. In imaging systems, for instance, imperceptible digital noise can be introduced to deliberately trigger high-confidence classification errors

during jamming operations. Notable variations include "white-box" attacks, where the adversary possesses complete knowledge of the deep neural network's internal mechanics, and "black-box" attacks, where the adversary is only aware of the system's inputs and outputs. To address these vulnerabilities, researchers at Stony Brook University (New York) and IBM engineered the ARES evaluation framework [36]. Based on reinforcement learning for adversarial ML, this framework enables the analysis of system-level attack and defense dynamics, allowing for a holistic re-evaluation of target defense frameworks.

- **Transparency:** Because they operate in safety-critical domains, military ML applications require high degrees of transparency, stringent security, and verifiable user trust. Addressing transparency challenges involves bolstering operator confidence in automated recommendations, uncovering novel causal links that can be validated via alternative means, defining operational boundaries, eliminating systemic biases to ensure equitable treatment across cases, and enhancing model interpretability. This allows operators to anticipate system outputs, comprehend model variables, and interpret the underlying training process.
- **Ethical dimensions of automated decision-making:** The transition to automated decision-making risks diminishing the "human element" of warfare. An autonomous system lacks the capacity to evaluate civilian casualties as collateral damage or weigh the moral implications of neutralizing an enemy combatant, even if that combatant indicates an intent to surrender. Teaching a machine to comprehend and integrate these ethical benchmarks into its algorithmic logic remains an immense hurdle.
- **Data scarcity and missing parameters:** The efficacy of an ML model depends fundamentally on sample quality, data volume, and the contextual relevance of the information to the operational environment. For example, despite the United States Army being a major global power with extensive combat history and vast archives of recorded data, its available datasets may still be insufficient for deploying ML in an active, complex, and immediate conflict. This stems from a lack of comprehensive data on adversaries, coupled with the immense difficulty of computing, processing, and extracting data in real-time during live combat. Furthermore, when training data originates purely from military exercises, it is inherently constrained by safety protocols and budgetary limits, making it fundamentally distinct from actual warfare. Proposed solutions include intensive fieldwork to gather, identify, and label relevant real-world parameters for database creation. Alternatively, data can be generated via real-time strategy simulations, allowing commanders to simulate diverse scenarios and accumulate tactical experience in digital formats.
- **Deficiencies in evaluation metrics:** The ultimate objective of military ML is to construct systems that enhance decision-making by leveraging historical data alongside insights gained from unfamiliar environments. However, establishing the boundary conditions of an algorithm's validity and ensuring its adaptability to novel scenarios remains highly challenging. Consequently, verifying the efficacy of these decision-making architectures requires extensive simulation testing and rigorous experimentation.
- **Tactical environment modeling complexity:** Capturing the full spectrum of battlefield information—including the status of active combat units and weapon systems—presents severe modeling difficulties. This complexity is compounded by the unpredictability of unit behaviors, the fluid nature of the combat environment, vague action mechanisms, and poorly defined evaluation benchmarks. Additionally, developers must account for variables such as wartime cost-efficiency, asset sustainability, structural damage, and the strategic necessity of establishing air, land, or maritime dominance.

Conversely, geopolitical and macroeconomic dynamics are also heavily at play, given that military targets are frequently the direct manifestation of political choices. Consequently, even when the immediate outcome of an engagement is sub-optimal, the overall strategic effect within a tactical environment can remain positive and valid if effective political and economic containment is successfully achieved. These compounding variables create immense friction when attempting to construct the architectural upper layers of the modeling system.

- **Information constraints and operational uncertainty:** During active combat operations, incoming intelligence is often fragmented, and the reliability of the originating source is frequently ambiguous. Consequently, executing commands under these restrictive conditions does not guarantee a net strategic advantage. To mitigate these obstacles, it is critical to determine optimal methods for discretizing time and operational actions, establishing temporary windows of opportunity, and seizing tactical initiative to achieve broader military or strategic goals.

Command structures that rely strictly on rigid, conventional routines or prescribed operational protocols risk finding themselves at a tactical disadvantage; instead, leadership should remain highly attentive to battlefield contingencies and actively innovate their methodologies as operational demands dictate.

Nevertheless, codifying the historical expertise of military professionals into structured facts and logical rules remains an indispensable prerequisite for launching the development of any ML algorithm designed for tactical environments. At a foundational level, the system must possess a rigorous understanding of the military domain itself, the performance parameters of deployed weapons systems, and various models of warfare (including asymmetric, symmetric, hybrid, conventional, and nuclear dimensions). It must also integrate pertinent decision-making frameworks (such as political alignment, economic stabilization, civilian protection, and territorial conquest), alongside established laws of armed conflict and rules of engagement between opposing forces or allied coalitions.

To contextualize these issues, the specific complications encountered during a naval war simulation designed to predict victorious warships using a Random Forest model [37] serve as a practical example. This case study, along with the associated training data challenges, is presented in **Table 2**.

Table 2. Example of a training data challenge.

Challenge	Random Forest-Based Battleship Outcome Prediction in Warfare Simulation [37]
Risk of friendly-fire incidents	This issue was not relevant in the study because the model was not designed to distinguish allied from hostile units; its sole purpose was to forecast the winning side.
Vulnerability to adversarial manipulation of ML systems	Opponents could potentially imitate the deployment of decoy weapons. It would therefore be necessary to investigate whether such tactics might distort the training dataset and influence prediction outcomes.
Transparency and explainability	For military decision-makers to fully rely on the model's forecasts, they must understand all underlying assumptions, data sources, and limitations. In this case, the inputs and methodology were relatively transparent, although the problem itself was simplified.
Ethical concerns regarding machine-generated outcomes	This challenge was not applicable because the model did not autonomously make operational decisions; it only estimated the likely winner of a battle.
Data scarcity and missing information	The research utilized 9,660 battleship records. Nevertheless, confidential factors—such as undisclosed weapons systems or emerging defensive technologies—were absent from the dataset.
Weaknesses in evaluation metrics	Due to the lack of actual battlefield observations and the constraints of the training data, it remained uncertain whether the model's predictions would be dependable and valid under real combat conditions.
Difficulty of representing the tactical environment	The analysis ignored contextual factors such as battlefield conditions and the initial deployment of forces. Instead, it relied only on ship characteristics including size, speed, capacity, crew count, attack power, additional attack capability, and defense strength.
Incomplete and uncertain information	The model produced a single predicted victor. To better reflect reality, time progression and operational actions would need to be discretized to account for events such as weapon malfunctions or ammunition limitations, which could substantially alter the outcome.

Conclusion

Machine learning (ML) has evolved into a ubiquitous catalyst for innovation, fundamentally restructuring diverse global industries, with the defense sector being no exception. The deployment of these technologies yields a broad spectrum of structural advantages, including heightened operational efficiency, substantial long-term cost reductions, highly sensitive fraud detection mechanisms, and the optimized oversight of intricate military supply networks. Concurrently, however, embedding ML within defense operations exposes complex legal and ethical dilemmas that command serious attention. At present, a distinct regulatory void persists regarding comprehensive legal and ethical frameworks capable of governing these emerging technologies. This policy deficit introduces immediate risks, as the absence of explicit, standardized guidelines could inadvertently facilitate the emergence of unforeseen operational hazards. Accordingly, the proactive formulation of an overarching legal architecture is of paramount importance to pre-emptively mitigate the latent vulnerabilities and systemic risks that ML integration might introduce into defense infrastructures.

Ethical considerations similarly exert a profound influence over the unchecked expansion of ML within defense frameworks, spanning both individual and institutional moral obligations. These ethical dilemmas become significantly more complex when applied to the preservation of human life and the protection of fundamental human rights that safeguard human dignity. Among the most heavily debated and contentious issues is the delegation of critical decision-making authority to autonomous systems driven by ML capabilities, which inherently operate without a capacity for moral discernment. While such delegation represents a significant break from historical military norms, it stands as a unique challenge demanding rigorous and earnest contemplation.

The intersection of these legal and ethical dynamics underscores the pressing need to carefully evaluate how AI and ML are applied within defense frameworks. International humanitarian laws must be strictly observed, and human rights must be vigilantly protected, requiring the establishment of unyielding ethical principles. Yet, despite widespread efforts to highlight the necessity of these governance frameworks, notable operational gaps remain, demonstrating that the trajectory toward actual implementation is inherently intricate and multifaceted.

Consequently, this study conducted a critical evaluation to delineate the core requirements and systemic vulnerabilities associated with the design of future legal and ethical governance frameworks. These frameworks must be carefully calibrated to account for the unique conditions of human nature that characterize the defense sector and its corresponding tactical scenarios. Throughout this analysis, a primary limitation emerged in the form of data scarcity, driven by the highly classified and confidential nature of the defense domain; this constraint, in turn, served as the central pillar for defining the boundaries of the proposed frameworks. While establishing standardized legal structures and resolving ethical tensions present significant hurdles, navigating these complexities is absolutely necessary to capture potential technological benefits while minimizing operational risks and collateral harms. Simultaneously, this assessment accounted for the distinct advantages of augmenting defense architectures with ML, such as the enhancement of combatant training simulation systems and the precise prediction of tactical situations to maximize mission success rates. Conversely, the identified disadvantages include the immense difficulty of executing robust testing phases for non-deterministic systems, alongside steep initial deployment barriers. Furthermore, the core challenges of applying ML to defense initiatives were analyzed and categorized, emphasizing the need for complexity reduction and accelerated recovery times following system failures. These persisting challenges and latent opportunities map out vital future research directions necessary to advance machine learning applications in defense while simultaneously addressing legal, ethical, and operational realities.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 07 Dec 2024

Revised: 25 Jan 2025

Accepted: 24 Mar 2025

Published online: 15 September 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

- Loubiri O, Maag S. Automated web testing using machine learning and containerization. In: Proceedings of the 2022 26th International Conference on Circuits, Systems, Communications and Computers (CSCC); 2022 Jul 19-22; Crete, Greece. Piscataway (NJ): IEEE; 2022. p.113-21.
<https://doi.org/10.1109/CSCC55931.2022.00029>.
- Dhakal A, Kulkarni SG, Ramakrishnan KK. ECML: improving efficiency of machine learning in edge clouds. In: Proceedings of the 2020 IEEE 9th International Conference on Cloud Networking (CloudNet); 2020; Piscataway (NJ). IEEE; 2020. p.1-6.
- Dhirawani P, Parekh R, Kandoi T, Srivastava K. Cost savings estimation for solar energy consumption using machine learning. In: Proceedings of the 2022 11th International Conference on Renewable Energy Research and Application (ICRERA); 2022; Istanbul, Turkey. IEEE; 2022. p.42-9.
- Liulys K. Machine learning application in predictive maintenance. In: Proceedings of the 2019 Open Conference of Electrical, Electronic and Information Sciences (eStream); 2019; Vilnius, Lithuania. IEEE; 2019. p.1-4.
- Wei Y, Qi Y, Ma Q, Liu Z, Shen C, Fang C. Fraud detection by machine learning. In: Proceedings of the 2020 2nd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI); 2020; Taiyuan, China. IEEE; 2020. p.101-15.
- Rao UA, Nalinipriya G, Jangirala S, Poonguzhali I, Azahad S, Samatha B. Improved food supply chain forecasting management using IoT and machine learning. In: Proceedings of the 2022 IEEE 2nd Mysore Sub Section International Conference (MysuruCon); 2022; Mysuru, India. IEEE; 2022. p.1-4.
- Sudharsan B, Kumar SP, Dhakshinamurthy R. AI Vision: smart speaker design and implementation with object detection custom skill and advanced voice interaction capability. In: Proceedings of the 2019 11th International Conference on Advanced Computing (ICoAC); 2019; Chennai, India. IEEE; 2019. p.97-102.

Ramadhan G, Setiawan EB. Collaborative filtering recommender system based on memory based in Twitter using decision tree learning classification (case study: movie on Netflix). In: Proceedings of the 2022 International Conference on Advanced Creative Networks and Intelligent Systems (ICACNIS); 2022; Bandung, Indonesia. IEEE; 2022. p.1-6.

Monzon Baeza V, Concha Salor L. New horizons in tactical communications: an overview of emerging technologies possibilities. *IEEE Potentials*. 2024;43(1):12-9.
<https://doi.org/10.1109/MPOT.2023.3297326>.

Concha Salor L, Monzon Baeza V. Harnessing the potential of emerging technologies to break down barriers in tactical communications. *Telecom*. 2023;4(4):709-31.
<https://doi.org/10.3390/telecom4040032>.

Villar Miguelez C, Monzon Baeza V, Parada R, Monzo C. Guidelines for renewal and securitization of a critical infrastructure based on IoT networks. *Smart Cities*. 2023;6(2):728-43.
<https://doi.org/10.3390/smartcities6020035>.

Parada R, Monzon Baeza V, Barraca-Ibort DN, Monzo C. LoRa-based low-cost nanosatellite for emerging communication networks in complex scenarios. *Aerospace*. 2023;10(9):754.

Petrov V, Lema MA, Gapeyenko M, Antonakoglou K, Moltchanov D, Sardis F, et al. Achieving end-to-end reliability of mission-critical traffic in softwarized 5G networks. *IEEE J Sel Areas Commun*. 2018;36:485-501.

Spantideas S, Giannopoulos A, Cambeiro MA, Trullols-Cruces O, Atxutegi E, Trakadas P. Intelligent mission critical services over beyond 5G networks: control loop and proactive overload detection. In: Proceedings of the 2023 International Conference on Smart Applications, Communications and Networking (SmartNets); 2023; Istanbul, Turkey. IEEE; 2023. p.1-6.

Skarin P, Tärneberg W, Årzen KE, Kihl M. Towards mission-critical control at the edge and over 5G. In: Proceedings of the 2018 IEEE International Conference on Edge Computing (EDGE); 2018; Washington (DC). IEEE; 2018. p.50-7.

Bongirwar V, Mishra SS, Bisen A, Mundhada S, Singh U. Disaster information verification and validation application using machine learning. *Int J Next Gener Comput*. 2022;13.

Arinta RR, Andi WR E. Natural disaster application on big data and machine learning: a review. In: Proceedings of the 2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE); 2019; Yogyakarta, Indonesia. IEEE; 2019. p.249-54.

Liu Y, Deng Y, Nallanathan A, Yuan J. Machine learning for 6G enhanced ultra-reliable and low-latency services. *IEEE Wirel Commun*. 2023;30:48-54.

Alcántara Suárez EJ. Análisis de la aplicación de machine learning en sistemas de defensa [master's dissertation]. Barcelona: Universitat Oberta de Catalunya; 2023.

Campbell RG, Eirinaki M, Park Y. A curriculum framework for autonomous network defense using multi-agent reinforcement learning. In: Proceedings of the 2023 Silicon Valley Cybersecurity Conference (SVCC); 2023; San Jose (CA). IEEE; 2023. p.1-8.

Fernández-Barrero D, Fontenla-Romero O, Lamas-López F, Novoa-Paradela D, R-Moreno MD, Sanz D. SOPRENE: assessment of the Spanish Armada's predictive maintenance tool for naval assets. *Appl Sci*. 2021;11(16):7322.
<https://doi.org/10.3390/app11167322>.

Kseniia N, Minbaleev A. Legal support of cybersecurity in the field of application of artificial intelligence technology. In: Proceedings of the 2020 International Conference Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS); 2020; Yaroslavl, Russia. IEEE; 2020. p.59-62.

Yongcui Z. Situation awareness and target recognition of marine big data battlefield based on deep learning. In: Proceedings of the 2022 IEEE International Conference on Artificial Intelligence and Computer Applications (ICAICA); 2022; Dalian, China. IEEE; 2022.

p.430-5.

Ergünay ÜSK, Kahraman F, Ates HF. Automated detection of refugee dwellings from satellite imagery using multi-class graph-cut segmentation and shadow information. In: Proceedings of the 2014 22nd Signal Processing and Communications Applications Conference (SIU); 2014; Trabzon, Turkey. IEEE; 2014. p.1591-5.

Korecki Z, Smrž V, Pecháček T. Support of the logistics chain by small aircraft in times of emergency. In: Proceedings of the 2020 New Trends in Aviation Development (NTAD); 2020; Sary Smokovec, Slovakia. IEEE; 2020. p.136-41.

European Commission. White paper on artificial intelligence: a European approach to excellence and trust [Internet]. Brussels: European Commission; 2020 [cited 2023 Jul 27].

European Commission. Ethics guidelines for trustworthy AI [Internet]. Brussels: European Commission; 2020 [cited 2023 Jul 27].

Truby J, Brown R, Ibrahim I, Parellada O. A sandbox approach to regulating high-risk artificial intelligence applications. *Eur J Risk Regul.* 2022;13:270-94.

UNESCO. Artificial intelligence: examples of ethical dilemmas [Internet]. Paris: UNESCO; 2023 [cited 2023 Jul 27].

Tarazona Lizarraga C. Análisis de las necesidades de una Smart City en el marco de un desarrollo sostenible [master's dissertation]. Barcelona: Universitat Oberta de Catalunya; 2020.

Wakabayashi D, Shane S. Google will not renew Pentagon contract that upset employees. *New York Times* [Internet]. 2018 [cited 2023 Jul 27].

Roldán Tudela JM. Artificial intelligence application in defence. Madrid: Spanish Institute of Technological Studies; 2018. p.166.

Breaking Defense. BD checks out Army's robotic gun: ATLAS 2020 [Internet]. 2020 [cited 2023 Jul 27].

Strout N. How the Army plans to revolutionize tanks with artificial intelligence. *C4ISRNET* [Internet]. 2020 [cited 2023 Jul 27].

Wang W, Liu H, Lin W, Chen Y, Yang JA. Investigation on works and military applications of artificial intelligence. *IEEE Access.* 2020;8:131614-25.

<https://doi.org/10.1109/ACCESS.2020.3009840>.

Ahmed F, Vaishnavi P, Eykholt K, Rahmati A. Ares: a system-oriented wargame framework for adversarial ML. In: Proceedings of the 43rd IEEE Symposium on Security and Privacy Workshops (SPW); 2022; San Francisco (CA). IEEE; 2022. p.73-9.

<https://doi.org/10.1109/SPW54247.2022.9833895>.

Intizhami NS, Husodo AY, Jatmiko W. Warfare simulation: predicting battleship winner using random forest. In: Proceedings of the 2019 IEEE International Conference on Communication, Networks and Satellite (Comnetsat); 2019; Makassar, Indonesia. IEEE; 2019. p.30-4.