13th International Conference on Sustainable Development, 10-11 September 2025, Rome, Italy
https://ecsdev.org

GOVERNING AI AGENTS IN NATIONAL DEFENSE: FROM STRATEGIC AUTONOMY TO ACCOUNTABLE DECISION ARCHITECTURES

Dumitru-Alexandru Bodislav,
Bucharest University of Economic Studies, Romania
alex.bodislav@ase.ro

Florina Bran
Bucharest University of Economic Studies, Romania
florina.bran@eam.ase.ro

Călin Bogdan Chiotan
University of Valachia, Romania
radu.chiotan@mk.ase.ro

Vicențiu Mihai Mateescu
Bucharest University of Economic Studies, Romania
mateescuvicentiu15@stud.ase.ro

Abstract

This paper examines the integration of artificial intelligence (AI) agents into national defense strategies, focusing on their role in strategic planning, intelligence analysis, logistics optimization, and decision support systems. Moving beyond the narrow debate on lethal autonomous weapons, the study conceptualizes AI agents as governance instruments that reshape command structures, civil–military relations, and strategic autonomy. Using a policy analysis framework, the paper explores how AI agents alter the speed, scale, and opacity of defense decision-making, generating both operational advantages and systemic risks. The analysis identifies three critical policy tensions: (i) autonomy versus human oversight in strategic and operational decisions, (ii) efficiency gains versus accountability and democratic control, and (iii) national security imperatives versus international norms and escalation risks. The paper proposes a layered governance model for defense AI agents, combining human-in-the-loop mandates, algorithmic auditability, and constitutional alignment of military AI use. The contribution lies in reframing AI agents not merely as technological assets, but as strategic actors embedded in institutional and legal architectures, requiring explicit policy design to preserve civilian control, strategic stability, and democratic legitimacy in an AI augmented defense ecosystem.

Keywords: Artificial intelligence agents; defense policy; military decision-making; strategic autonomy; civil–military relations; national security strategy; defense innovation

Introduction: AI Agents and the Transformation of Defense Governance

In the last ten years, the use of artificial intelligence (AI) in national defense systems has sped up a lot (Burlacu et al., 2025; Rădulescu et al., 2025). This is in line with bigger changes in digital technologies and data-driven decision-making. Military organizations are using more and more computer systems that can handle a lot of data, find patterns in complicated situations, and help with both operational and strategic planning. In this changing world of technology, AI agents—computers that can see their surroundings, process information, and act toward set goals—have become important parts of modern defense systems (Burlacu, Crețu et al., 2024). They can be used for a wide range of things, including intelligence analysis, cyber defense, logistics coordination, surveillance systems, and decision-support tools that help commanders and policymakers. As these technologies become more advanced and widespread, they are starting to have an effect on both battlefield capabilities and the way defense governance is set up (Russell and Norvig, 2021).

Literature Review: Artificial Intelligence, Military Autonomy, and Governance

The swift proliferation of artificial intelligence technologies within defense institutions has led to an expanding corpus of scholarly literature investigating the strategic, technological, and ethical ramifications of algorithmic systems in military settings. Current research predominantly revolves around three interconnected debates: the strategic function of AI in military rivalry, the ethical and legal dilemmas presented by autonomous weapon systems, and the nascent domain of algorithmic governance and institutional accountability. Although these literary strands offer significant insights into the technological and normative ramifications of defense AI, they frequently tackle these aspects in isolation. Consequently, the extensive governance ramifications of AI agents integrated within military decision-making frameworks are still relatively unexamined.

AI Agents as Governance Instruments in Defense Decision Architectures

The growing use of artificial intelligence in national defense systems needs a conceptual framework that can cover both the technological capabilities of AI agents and the effects they have on institutions. Conventional methodologies typically regard artificial intelligence as a functional instrument aimed at improving military efficacy; however, this viewpoint is inadequate for comprehending the transformative impact of AI systems on the structure of defense decision-making. AI agents are becoming more common in areas like strategic planning, intelligence analysis, logistics coordination, and cyber defense, where they change how information is processed and how decisions are made. Consequently, AI systems operate as mediating structures integrated within institutional processes rather than merely as technical tools (Russell and Norvig, 2021).

Governance Tension in AI-Augmented Defense Systems

The incorporation of artificial intelligence agents into national defense infrastructures creates substantial operational prospects while simultaneously presenting intricate governance dilemmas. As algorithmic systems become a part of intelligence analysis, logistics coordination, cyber defense, and strategic decision support, they change the way military decisions are made and carried out. AI systems can make things much more efficient, faster, and able to predict the future, but using them also causes problems between automation and oversight, technological performance and democratic accountability, and national security goals and international stability. To make sure that AI-enabled defense systems stay in line with democratic norms and strategic stability, we need to understand these tensions (Horowitz, 2018).

Toward an Accountable Governance Architecture for Defense AI Agents

The previous analysis shows how adding AI agents to national defense systems creates structural tensions between operational efficiency, institutional accountability, and strategic stability. AI technologies offer substantial benefits in fields like intelligence analysis, predictive threat detection, and logistics optimization. However, their increasing role in defense decision-making necessitates governance frameworks that can align technological efficacy with democratic oversight and legal accountability. To deal with these problems, we need to go beyond reactive regulation and instead create institutional structures that include governance principles directly into the use and deployment of defense AI systems (Floridi et al., 2018).

Conclusion

The integration of artificial intelligence agents into national defense systems represents one of the most significant technological and institutional transformations in contemporary security governance. As AI technologies increasingly permeate intelligence analysis, logistics management, cyber defense, and strategic planning, their influence extends far beyond traditional battlefield applications. Rather than functioning solely as operational tools, AI agents are becoming embedded within the institutional architectures that structure how defense decisions are generated, evaluated, and implemented. This transformation introduces important opportunities for improving the analytical capacity and responsiveness of military institutions, but it also raises complex governance challenges concerning accountability, transparency, and strategic stability.

This study has argued that understanding the implications of artificial intelligence in defense contexts requires a conceptual shift from viewing AI primarily as a technological capability toward interpreting it as a governance instrument embedded within defense decision architectures. From this perspective, AI agents function as mediating structures that influence the flow of information, shape threat assessments, and affect the range of strategic options considered by human decision-makers. As a result, the integration of algorithmic systems alters the institutional dynamics of military command and introduces new forms of interaction between human authorities and automated analytical processes.

The analysis identified three central governance tensions emerging from the adoption of AI agents within defense institutions. First, the increasing autonomy of algorithmic systems may challenge the preservation of meaningful human oversight, particularly in high-speed operational environments where decision cycles are compressed by automated analysis and predictive modeling. Second, while AI technologies can significantly improve efficiency and operational effectiveness, their complexity and opacity may weaken traditional mechanisms of democratic accountability and civilian oversight. Third, the growing strategic competition among states to develop advanced AI capabilities risks accelerating technological deployment before adequate governance frameworks are established, potentially destabilizing international security environments and increasing the likelihood of unintended escalation.

To address these tensions, the paper proposed a layered governance architecture for defense AI systems that integrates three core principles: human-in-the-loop decision structures, algorithmic auditability, and institutional alignment with constitutional frameworks governing civilian control over the military. By embedding these governance mechanisms within defense decision architectures, states can harness the operational advantages of AI technologies while preserving the institutional safeguards necessary for democratic legitimacy and strategic stability. This approach emphasizes that the effective governance of defense AI systems must be incorporated directly into institutional design rather than treated as an external regulatory constraint applied after technological deployment.

Despite its conceptual contributions, this study also presents several limitations that should be acknowledged. First, the analysis adopts a primarily conceptual and policy-oriented approach rather than relying on empirical case studies of specific defense institutions. Due to the sensitive nature of military technologies and the limited public availability of detailed information regarding AI deployments in national defense systems, empirical data on operational practices remains relatively scarce. As a result, the arguments presented in this study focus on theoretical and institutional analysis rather than on systematic empirical evaluation of existing AI-enabled defense programs.

Second, the paper focuses primarily on governance challenges within democratic political systems, where civilian oversight and institutional accountability represent central normative principles of defense governance. While this perspective is appropriate for examining the relationship between AI technologies and democratic legitimacy, it may not fully capture the governance dynamics present in authoritarian or hybrid political systems, where the relationship between military institutions, political leadership, and technological governance may differ significantly. Comparative research across different political systems may therefore provide valuable insights into how AI governance frameworks evolve under diverse institutional conditions.

A third limitation concerns the rapidly evolving nature of artificial intelligence technologies themselves. Advances in machine learning, autonomous systems, and data analytics continue to expand the capabilities of AI agents at a pace that may outstrip the development of governance frameworks designed to regulate their use. As new forms of AI-enabled decision support and autonomous systems emerge, the governance challenges discussed in this paper may evolve in ways that require continuous institutional adaptation and policy innovation.