European Journal of Sustainable Development (2026), **15**, 4, 197-210 ISSN: 2239-5938
Doi: 10.14207/ejsd.2026.v15n4p197

Governing AI Agents in National Defense: From Strategic Autonomy to Accountable Decision Architectures

By Dumitru-Alexandru Bodislav*[1], Florina Bran[1], Călin Bogdan Chiotan[2], Vicenţiu Mihai Mateescu[1]

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

1. 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 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, Creţu 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).

[1]Bucharest University of Economic Studies, Romania
[2] University of Valachia, Romania·

Historically, discussions concerning artificial intelligence in military settings have predominantly concentrated on the ethical and legal ramifications of Lethal Autonomous Weapons Systems (LAWS). A lot of the research that is already out there looks at the dangers of giving machines the power to make deadly decisions. This raises questions about who is responsible, how to hold people accountable, and whether autonomous weapons can be used in accordance with international humanitarian law. These debates are still important, but they only show a small part of the bigger change that is happening in defense institutions right now. AI agents are becoming more common in the early stages of the defense decision cycle, such as interpreting intelligence, finding threats, using predictive analytics, and making strategic plans. In these scenarios, AI systems operate not as independent weapons but as decision-support frameworks that influence the interpretation of information and the formulation of responses to emerging security threats by military entities and political leaders (Scharre, 2018).

The increasing dependence on AI-driven decision support systems brings about substantial transformations in the structure and dynamics of military command. In the past, defense organizations used hierarchical decision-making systems where people look at intelligence inputs, talk about different strategies, and then give the go-ahead for operations. AI agents add another layer of algorithmic mediation to these processes. AI systems can greatly speed up decision-making by filtering large datasets, making probabilistic forecasts, and suggesting the best course of action. In areas where reaction time is very important, like cyber warfare, missile defense, and real-time battlefield analytics, this kind of speedup can give you an edge. But the fact that algorithmic decision systems are getting faster and more complicated also raises questions about how clear they are, how easy they are to understand, and how human judgment might be pushed aside in important strategic decisions (Horowitz, 2018).

In addition to operational concerns, incorporating AI agents into defense decision-making processes has wider effects on how institutions are run. When algorithmic systems become a big part of strategic planning and intelligence analysis, they change the way military groups, political leaders, and civilian oversight groups work together. In democracies, the idea that civilians should oversee the military means that defense decisions must be made with the approval of elected officials and be subject to institutional checks and balances (Burlacu et al., 2023; Rădulescu, Mănescu et al., 2023). But algorithmic decision systems often work with complicated models that policymakers and oversight bodies may not be able to understand or judge. The growing lack of transparency in AI-driven decision-making could lead to technical experts or automated analytical systems having more power, which could make it harder for traditional democratic accountability systems to work (Brundage et al., 2018).

The swift integration of artificial intelligence into defense systems must be comprehended within a wider geopolitical framework marked by escalating technological rivalry among dominant nations. Governments are starting to see AI as a strategic tool that could change the balance of power in the military and in the world. So, many countries have started ambitious national plans to speed up the development of AI and include these technologies in their defense planning and security systems. While these kinds of projects may improve operational and intelligence capabilities, they also run the risk of encouraging the quick deployment of AI systems before strong governance frameworks are in place.

This dynamic may heighten the probability of inadvertent escalation, misinterpretation of algorithmically generated threat assessments, or destabilizing interactions among technologically advanced military systems (Allen and Chan, 2017).

Given these changes, we need to go beyond the usual idea that artificial intelligence is just a way to make military technology better. Instead, AI agents should be seen as parts of institutions that are built into larger systems of governance that affect how defense decisions are made, judged, and carried out. From this standpoint, the primary challenge confronting modern defense institutions is not merely the adoption of AI systems, but the governance of their integration in a manner that maintains strategic stability, democratic legitimacy, and efficient civilian oversight. To meet this challenge, we need governance frameworks that can find a balance between the operational benefits of algorithmic decision support and the moral requirements of accountability, transparency, and political responsibility (Floridi et al., 2018).

This paper tackles these problems by thinking of AI agents as tools for governance in national defense systems. From a policy analysis standpoint, it investigates the impact of AI integration on defense decision-making frameworks and elucidates the principal tensions arising among strategic autonomy, accountability, and international stability. The research subsequently introduces a stratified governance framework aimed at harmonizing operational efficiency with democratic oversight, highlighting mechanisms such as human-in-the-loop decision mandates, algorithmic auditability, and constitutional conformity in military AI deployment. By redefining AI agents as strategic entities integrated within institutional and legal frameworks rather than merely technological instruments, the paper enhances the existing literature on AI governance and underscores the significance of intentional policy formulation in addressing the enduring consequences of AI-enhanced defense ecosystems.

2. 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.

The initial body of research examines the strategic ramifications of artificial intelligence within the framework of international security competition. Scholars have increasingly contended that AI constitutes a general-purpose technology with the potential to alter the equilibrium of military power in manners akin to previous technological revolutions, including nuclear weapons, precision-guided munitions, and information warfare capabilities (Nica et al., 2023). From this point of view, AI makes it easier for countries to handle large amounts of data, find strategic patterns, and plan complicated

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

military operations across many areas. Defense analysts have stressed that AI-enabled systems can greatly enhance intelligence gathering, predictive threat assessment, cyber defense capabilities, and operational logistics, making modern armed forces more efficient and responsive (Horowitz, 2018). The strategic importance of AI has prompted major powers to invest significantly in national AI strategies that emphasize both civilian innovation and defense applications, indicative of an increasing belief that technological supremacy in artificial intelligence could affect future geopolitical dynamics (Allen and Chan, 2017).

Researchers have also looked at how AI technologies affect the speed and flow of military decision-making within this strategic framework. By adding machine learning models and predictive analytics to intelligence systems, defense organizations can find patterns in huge datasets that would be hard for human analysts to find in real time. This ability can speed up the pace of military decision-making, which could shorten the time between finding a threat and coming up with a plan to deal with it. However, researchers have also said that speeding things up like this could create new risks, especially if automated systems make suggestions that decision-makers feel they must follow without thinking them through carefully. The term "hyperwar" which comes from the field of defense technology, shows this worry by saying that AI-enabled military systems could make conflict move at machine-speed decision cycles that make it hard for people to command (Scharre, 2018).

A second large group of writings looks at the moral and legal issues that come up with autonomous weapons systems. These conversations have been especially important in debates about Lethal Autonomous Weapons Systems (LAWS), where scholars and policymakers have looked into the possibility that AI-enabled weapons could choose and attack targets without any help from humans. Critics of these systems contend that entrusting lethal decisions to machines engenders significant ethical dilemmas and may compromise fundamental tenets of international humanitarian law, such as the principles of distinction, proportionality, and accountability in the application of force (Asaro, 2012). The primary focus of this literature is the possible development of a "responsibility gap," which complicates the assignment of legal or moral accountability for the actions of autonomous systems (Sharkey, 2019).

Considering these apprehensions, numerous scholars and international entities have promoted the preservation of "meaningful human control" over crucial military decisions pertaining to the application of force. This idea has become a guiding principle in ongoing international talks about how to control autonomous weapons systems, especially in the context of the United Nations Convention on Certain Conventional Weapons. The goal of keeping meaningful human control is to make sure that human judgment is still the most important factor in decisions about using deadly force. This way, military operations can still be held accountable and morally responsible (Scharre, 2018). Nonetheless, the emphasis on lethal autonomy has constrained the discourse by concentrating predominantly on weapon systems rather than on the comprehensive integration of AI into defense decision-making frameworks.

A third and more relevant area of research looks at artificial intelligence in terms of governance and institutional accountability. Studies in algorithmic governance demonstrate that AI systems can transform decision-making processes in public and

private institutions by implementing automated analytical frameworks that regulate information access and organize policy options. These studies underscore that algorithmic systems are not impartial technical instruments; instead, they are socio-technical constructs situated within institutional frameworks that influence the processes of decision-making and justification (Kroll et al., 2017). In the public sector, researchers have emphasized the necessity of transparency, explainability, and auditability in algorithmic systems to uphold democratic oversight and institutional legitimacy.

From a governance standpoint, the growing integration of AI in defense institutions prompts significant inquiries concerning transparency, accountability, and institutional oversight. Military organizations frequently function in environments of secrecy that restrict external scrutiny, and the incorporation of intricate algorithmic systems may further obfuscate the decision-making processes that inform strategic and operational choices. If policymakers, legislators, or oversight bodies are unable to adequately assess the performance of AI-driven decision systems, the capacity of democratic institutions to exert substantial control over military operations may be compromised. This apprehension has prompted scholars to contend that governance frameworks for AI must transcend mere technical safety standards to incorporate institutional mechanisms that guarantee accountability and political responsibility (Floridi et al., 2018).

Even though these existing research streams have made important contributions, there is still a big conceptual gap. A lot of the writing about AI either sees it as a technological advancement that makes the military more effective or as a moral problem related to autonomous weapons. There has been limited focus on the role of AI agents as institutional actors integrated within defense governance frameworks. When AI systems sort through intelligence, make strategic predictions, or suggest operational choices, they change how information moves through military hierarchies and how strategic choices are finally made. In this regard, AI agents not only assist defense operations; they also transform the decision-making frameworks that underpin defense governance (Burlacu, Pargaru et al., 2022).

To fill this gap, we need to change the way we think about artificial intelligence from seeing it as just a technological tool to seeing it as a tool for governance that is built into institutional structures. This viewpoint underscores the necessity to investigate the impact of AI agents on the allocation of authority, the clarity of decision-making processes, and the frameworks of accountability within national defense systems. Consequently, the subsequent section formulates a conceptual framework that regards AI agents as institutional intermediaries within defense decision architectures, establishing a foundation for the examination of governance challenges linked to AI-enhanced military systems.

3. 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

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

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).

From this broader point of view, AI agents can be thought of as tools for governance that change how information flows, how power is shared, and how decisions are made in military organizations. In this case, governance means the rules and procedures that defense institutions and other complex organizations use to make, check, and carry out decisions. When AI agents are integrated into these processes, they alter the dynamics of information interpretation and the formulation of strategic decisions. To comprehend AI governance in defense, it is essential to examine the impact of algorithmic systems on the dynamics among human decision-makers, institutional hierarchies, and operational contexts (Kroll et al., 2017).

The first aspect of this conceptual framework pertains to the function of AI agents as decision facilitators within military command structures. Artificial intelligence systems greatly increase the ability of defense organizations to handle large amounts of data and make predictions. Machine learning algorithms can find patterns in intelligence datasets, satellite images, cyber traffic, and battlefield telemetry much faster than people can. With this ability, defense organizations can find threats sooner, predict how enemies will act, and improve their responses to situations. But speeding up decision-making cycles may also put pressure on command structures to become more automated, especially in fast-paced operational settings like cyber warfare or missile defense systems. As decision timelines shorten, human operators may find it increasingly challenging to assess algorithmic recommendations prior to the execution of operational actions, prompting significant concerns regarding the maintenance of substantive human oversight over strategic decisions (Horowitz, 2018).

A second aspect pertains to the function of AI agents as institutional intermediaries of information within defense entities. Military decision-making relies heavily on how intelligence data is interpreted. AI systems are increasingly important for filtering, organizing, and prioritizing the information that human analysts and commanders see. AI agents can find patterns and make threat assessments that change how people see strategic environments by using predictive analytics and probabilistic modeling. In this way, algorithmic systems affect not only how quickly decisions are made, but also how people think about security threats and what their operational priorities are. AI systems depend on training data and built-in optimization criteria, which means they might add systematic biases or structural assumptions to strategic assessments. As a result, the outputs produced by AI agents can affect how decision-makers think about strategy in ways that are hard to see without clear and auditable algorithmic systems (Floridi et al., 2018).

A third aspect of the conceptual framework pertains to the evolving function of AI agents as quasi-actors within defense decision architectures. Even though AI systems don't have legal agencies or their own intentions, they can have an effect on how decisions are made in institutions that is similar to some types of strategic agency. AI systems influence the choices that people can make when they make predictive forecasts, suggest

operational strategies, or automate logistical coordination. In complicated operational settings where algorithmic suggestions are relied upon for their analytical precision, the results generated by AI agents can have a substantial impact on ultimate choices. AI systems can indirectly influence strategic outcomes, even when human agents maintain formal decision-making authority. This occurrence underscores the necessity of perceiving AI not solely as a technological instrument but as an integral element within governance frameworks (Scharre, 2018).

This idea of AI agents as tools for governance shows that we need to go beyond just talking about how well algorithms work and instead focus on how AI-enabled decision systems are set up in institutions. Defense institutions must therefore contemplate the impact of AI integration on the equilibrium between automation and human supervision, the effect of algorithmic systems on the transparency of strategic decision-making, and the preservation of accountability within intricate human–machine interaction frameworks. These challenges are especially important in democratic political systems where military decision-making must answer civilian authorities and be checked by the constitution.

This conceptual framework posits AI agents as institutional mediators within defense governance structures, establishing a basis for the examination of systemic tensions arising from the integration of algorithmic decision systems into national security contexts. The growing dependence on AI-generated intelligence and decision support prompts essential inquiries regarding the interplay between operational efficiency and democratic accountability, the safeguarding of human judgment in strategic decision-making, and the alignment of AI-enabled defense systems with international norms regulating the use of force. The next part looks at the main governance tensions that come up in AI-enhanced defense systems. It looks at the balance between autonomy and oversight, efficiency and accountability, and national security needs and international stability.

4. Governance Tension in AI-Augmented Defense Systems

The incorporation of artificial intelligence agents into national defense infrastructure 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, 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).

The first big problem is that operational autonomy is growing, but human oversight is still needed in strategic and operational decision-making. AI agents can look at large amounts of data, find patterns, and make suggestions that can help or even automate some defense tasks. In areas like cyber defense, missile interception systems, and battlefield analytics, quick reaction times are very important. Algorithmic systems can make response times much faster than systems that are only run by people. But as automation becomes more common, the role of people who make decisions may slowly

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

change from direct control to oversight. This dynamic makes people worry that meaningful human control will be lost when people trust algorithmic outputs for their accuracy and efficiency. If decision-makers depend too much on automated suggestions, it may be harder for human operators to step in, especially in fast-paced work settings where response times are very short (Scharre, 2018).

The danger of too much automation goes beyond just operational mistakes; it also has bigger strategic effects (Bran et al., 2023; Rădulescu et al., 2022). AI systems use historical datasets and statistical models to make predictions and suggestions. These models can make some analytical tasks more accurate, but if the training data is incomplete, out of date, or not representative of new threats, they could also cause biases or misunderstandings. In military situations, these kinds of mistakes can lead to wrong threat assessments or wrong responses to escalation. Algorithmic decision systems frequently function as intricate "black boxes," making it challenging to pinpoint the exact origin of errors, thereby complicating operational accountability and strategic risk management (Russell and Norvig, 2021).

The second tension in governance comes from the fact that AI-enabled systems can make things more efficient, but democratic accountability and institutional oversight are still needed. To make their operations more efficient and make better use of their resources, defense organizations often use cutting-edge technology. AI systems can make logistics planning better, make intelligence analysis better, and make administrative work easier by automating routine analytical tasks. These abilities can make military organizations work better overall and let decision-makers react more quickly to changing security threats. Nonetheless, the opacity inherent in intricate machine learning systems may hinder the capacity of policymakers, regulatory entities, and the public to comprehend the formulation of strategic decisions when algorithmic systems are integral to decision support (Kroll et al., 2017).

In democratic political systems, where civilian authorities oversee keeping an eye on military institutions, the problem of algorithmic opacity is especially important. When defense strategies depend on analytical models that are hard to understand or check, oversight bodies like parliamentary committees or independent auditors may have a hard time figuring out if AI-generated suggestions are valid and trustworthy. In these situations, decision-making power may move to technical experts who oversee creating and keeping algorithmic systems running. This could make it harder for civilians to control military decisions in the usual ways. To keep democracy accountable, AI-enabled defense systems need to have governance mechanisms that make sure they are open, understandable, and traceable (Floridi et al., 2018).

A third governance tension pertains to the interplay between national security demands and the stability of international norms regulating the application of emerging technologies in warfare. As AI becomes more important to military modernization plans, countries may feel the need to speed up the use of AI-enabled systems to keep up with their enemies. This situation could lead to technological arms races, where AI systems are adopted too quickly without enough thought given to governance safeguards or international coordination. In this kind of situation, algorithmic decision systems could unintentionally make things worse by making it more likely that things will get worse if

automated threat detection systems misread signals or give false positives when there is geopolitical tension (Allen and Chan, 2017).

Furthermore, the implementation of AI-enhanced military systems prompts significant inquiries about the alignment of algorithmic decision-making with international humanitarian law and the established conventions regulating armed conflict. Even if AI systems are only used to help make decisions and not fully autonomous weapons, their outputs can still affect how targets are chosen, threats are assessed, and strategies are chosen in ways that make it harder to follow legal rules like proportionality and distinction. Without strong governance systems, the growing use of AI-generated analysis could make it harder to figure out who is responsible for military decisions, which could make it harder to enforce international legal norms (Asaro, 2012).

These three governance tensions—between autonomy and human oversight, efficiency and democratic accountability, and national security imperatives and international stability—show how hard it is to add AI agents to defense decision systems. To deal with these tensions, governance frameworks need to go beyond just technical regulation and include institutional mechanisms that promote transparency, accountability, and strategic stability. For example, defense institutions need to create systems that keep people in charge of important decisions, make sure that algorithmic systems can be checked and explained, and make sure that the use of military AI technologies is in line with constitutional and international law.

Considering these challenges, the next section suggests a layered governance model for AI agents in national defense systems. This framework aims to reconcile the operational benefits of AI-driven decision support with the normative imperatives of democratic accountability and strategic stability by incorporating human oversight mechanisms, algorithmic transparency standards, and institutional safeguards into defense governance structures.

5. Toward 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).

One of the most important ideas behind these kinds of governance structures is that people should still have control over important strategic decisions. AI agents can greatly enhance analytical capabilities and expedite information processing; however, they ought to function primarily as decision-support tools rather than autonomous decision-makers. Human-in-the-loop frameworks offer an institutional structure that ensures human operators retain accountability for sanctioning actions with strategic, political, or legal ramifications. In this setup, AI systems make analytical insights and probabilistic

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

predictions, but human decision-makers are still in charge of judging algorithmic suggestions and deciding what to do next. Keeping this structure is important for moral reasons, but it's also important for keeping military command systems accountable and making sure that political leaders are still in charge of making strategic decisions about the use of force (Scharre, 2018).

Defense AI agents need more than just human oversight; they also need systems that make sure the algorithms are clear and can be checked. Many modern machine learning models use complicated statistical structures that are hard to understand without specialized technical knowledge. This kind of complexity may make predictions more accurate, but it also runs the risk of making decision-making processes less clear, which hurts accountability in institutions. To address this issue, defense AI systems ought to integrate traceability mechanisms that can document the generation of algorithmic outputs and identify the data inputs that influence specific analytical conclusions. These mechanisms enable oversight institutions to reconstruct decision-making processes and assess the compliance of algorithmic systems with legal and institutional standards. Algorithmic auditability is thus an essential element of responsible AI governance in defense institutions (Kroll et al., 2017).

Another requirement for governance is that AI deployment must be in line with constitutional principles and democratic oversight structures. In democratic political systems, the validity of military decision-making is based on the principle of civilian oversight of the armed forces. As AI agents become more a part of intelligence analysis and operational planning, this principle needs governance systems that make sure algorithmic systems are still under political and institutional control. This might mean setting up specialized oversight groups that can look at AI-enabled defense systems, adding algorithmic risk assessments to the process of buying defense equipment, and making sure that military leaders and policymakers know enough about technology to judge AI-generated suggestions. These kinds of institutional arrangements can help keep decision-making power from moving too far into purely technical areas and make sure that strategic decisions stay part of democratic governance systems (Horowitz, 2018).

At the global level, the management of defense AI systems must also consider the possible destabilizing consequences of swift technological rivalry. As countries try to make and use more advanced AI, the lack of shared rules for governance could make it more likely that things will be misinterpreted, escalate, or lead to conflicts that weren't meant to happen. Setting up international rules for how to use AI responsibly in the military could help keep peace by making clear what is and isn't acceptable and encouraging openness about how AI-enabled systems are used. It may be hard to get all countries to agree on global rules in a competitive geopolitical environment, but measures to build trust, initiatives to make things more open, and frameworks for multilateral dialogue can all help lower the risks of algorithmic arms races (Allen and Chan, 2017).

These governance mechanisms work together to make a layered decision-making system for defense AI systems. This kind of architecture includes human oversight, algorithmic transparency, and institutional accountability in the design of AI-enabled military systems. Instead of treating governance as an outside rule that comes into play after technology is put to use, this method builds governance rules right into the way AI-assisted decision-making works. By doing this, defense organizations can use the analytical

benefits of AI while keeping the institutional protections that are needed for democratic legitimacy and strategic stability.

So, it is important to create accountable governance frameworks to make sure that AI agents help, not hurt, national defense institutions. When properly managed, AI systems can help military organizations analyze data better, be more aware of their surroundings, and make better strategic decisions. However, without careful planning for governance, the same technologies could make things less clear, take decision-making power away from people who are responsible, and make the risks of quick algorithmic decision cycles even higher. The challenge for modern defense organizations is not just to adopt new technologies, but also to create governance frameworks that can include AI agents in defense decision-making processes in a way that is compatible with democratic oversight, legal accountability, and international security stability.

6. 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 architecture 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.

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

To address these tensions, the paper proposed a layered governance architecture for defense AI systems that integrate 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.

Future research should therefore pursue several complementary directions in order to deepen our understanding of AI governance in defense contexts. One promising avenue involves comparative empirical analysis of national defense strategies to examine how different countries integrate AI technologies within their military institutions and what governance mechanisms they adopt to manage associated risks. Such comparative studies could illuminate the institutional diversity of defense AI governance models and identify best practices for balancing technological innovation with democratic accountability.

A second research direction concerns the development of institutional mechanisms for algorithmic transparency and explainability within military contexts. While significant progress has been made in civilian domains such as finance, healthcare, and

public administration, the application of explainable AI principles to defense systems remains relatively underexplored. Research focusing on the design of auditable and interpretable AI architectures for defense decision support could help bridge the gap between technical innovation and governance requirements.

Finally, future scholarships should examine the international governance dimensions of defense AI technologies. As artificial intelligence becomes an increasingly central component of military competition, the development of international norms, confidence-building measures, and cooperative governance frameworks will play an important role in mitigating the risks associated with algorithmic arms races. Interdisciplinary research combining international relations, technology policy, and legal analysis could contribute to the development of governance frameworks capable of promoting strategic stability while allowing responsible technological innovation.

In conclusion, the integration of AI agents into national defense systems represents both a technological and institutional transformation that will shape the future of security governance. Whether these technologies ultimately strengthen or undermine defense institutions will depend largely on the governance frameworks developed to regulate their deployment and operation. By recognizing AI agents as components of decision architecture rather than merely as technological tools, policymakers and scholars can better address the institutional challenges posed by AI-enabled defense systems and ensure that technological progress remains aligned with democratic principles, legal accountability, and international security stability.

References

Allen, G. and Chan, T. (2017) Artificial Intelligence and National Security. Cambridge, MA: Belfer Center for Science and International Affairs, Harvard Kennedy School.

Asaro, P. (2012) 'On banning autonomous weapon systems: Human rights, automation, and the dehumanization of lethal decision-making', International Review of the Red Cross, 94(886), pp. 687–709.

Bran, F., Bălu, O. F., Reynard, C., Rădulescu, C. V., Burlacu, S., Mănescu, C. O., & Gole, I. (2023). Mapping risk and performance business drivers post-pandemic: Future challenges. *Economic Computation & Economic Cybernetics Studies & Research*, 57(4).

Brundage, M. et al. (2018) The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. Oxford: Future of Humanity Institute, University of Oxford.

Burlacu, S., Bran, F., Rădulescu, C. V., & Mateescu, V. M. (2025). *The challenges of digitization in agriculture for sustainable rural development*. In Proceedings of the International Scientific Conference Sustainable Agriculture and Rural Development (pp. 963–977).

Burlacu, S., Creţu, R. C., Potcovaru, A. M. B., & Călin, A. M. (2024). Integrative digital governance: The role of public administration in promoting sustainable development amid post-pandemic technological innovations. *Proceedings of the International Conference on Business Excellence*, 18(1), 725–732.

Burlacu, S., Pargaru, I., Iacob, O. C., & Gombos, S. P. (2022). Digital public administration and the perspectives of sustainable development in Romania. *European Journal of Sustainable Development*, 11(4), 230–230.

Burlacu, S., Rădulescu, C. V., Bălu, E. P., & Dobre, F. (2023). Socio-cultural sustainability for resilience in public services: Strategies for achieving inclusive and equitable service delivery. *European Journal of Sustainable Development*, 12(3), 29–29.

Floridi, L. et al. (2018) 'AI4People—An ethical framework for a good AI society', Minds and Machines, 28(4), pp. 689–707.

Horowitz, M. C. (2018) 'Artificial intelligence, international competition, and the balance of power', Texas National Security Review, 1(3), pp. 36–57.

© 2026 The Authors. Journal Compilation © 2026 European Center of Sustainable Development.

Kroll, J. A. et al. (2017) 'Accountable algorithms', University of Pennsylvania Law Review, 165(3), pp. 633–705.

Nica, E., Poliak, M., Alpopi, C., Kliestik, T., Manole, C., & Burlacu, S. (2023). Impact of trade, FDI, and urbanization on female employment system in SAARC: GMM and quantile regression approach. *Systems*, 11(3), 137. https://doi.org/10.3390/systems11030137

Rădulescu, C. V., Bran, F., Ciuvăţ, A. L., Bodislav, D. A., Buzoianu, O. C., Ştefănescu, M., & Burlacu, S. (2022). Decoupling the economic development from resource consumption: Implications and challenges in assessing the evolution of forest area in Romania. *Land*, 11(7), 1097.

Rădulescu, C. V., Gâf-Deac, I. I., Bran, F., Burlacu, S., Bodislav, D. A., & Buzoianu, O. C. A. (2025). Cybernetic contributions of artificial intelligence to economic development in Romania. *Economic Computation & Economic Cybernetics Studies & Research*, 59(1).

Rădulescu, C. V., Mănescu, C. O., Popescu, M. L., & Burlacu, S. (2023). Sustainable development in public administration: Research, practice, and education. *European Journal of Sustainable Development*, 12(4), 27–27.

Russell, S. and Norvig, P. (2021) Artificial Intelligence: A Modern Approach. 4th edn. Pearson.

Scharre, P. (2018) Army of None: Autonomous Weapons and the Future of War. New York: W. W. Norton & Company.