Risks and Control Measures for Assuring the Safety of Trustworthy Autonomous Weapon Systems

Clara Maathuis[1] and Kasper Cools[2]
[1]Open University of the Netherlands, Heerlen, The Netherlands
[2]Belgian Royal Military Academy, Brussel, Belgium

clara.maathuis@ou.nl
kasper.cools@mil.be

Abstract: Autonomous Weapons Systems (AWS) represent a significant advancement in the military domain, offering potential benefits in precision, speed, and reduced human casualties, while simultaneously raising critical concerns regarding safety, ethics, and international security implications and consequences. While previous studies have extensively explored the technical, legal, and ethical aspects of AWS, there is a notable gap in addressing safety through the lens of building AWS as trustworthy systems. This article aims to bridge this gap by presenting a systematic analysis of the safety challenges associated with AWS and proposing robust control measures to address these concerns from a trustworthiness perspective. On this behalf, this research critically examines the inherent socio-technical risks of AWS, including potential system malfunctions, unintended engagements, ethical decision-making failures, and vulnerability to cyber attacks, evaluating these risks in the context of their potential impacts on combatants, civilians, and global stability. In response to these identified risks, a range of control measures designed to assure and enhance AWS safety, including advanced fail-safe mechanisms, multi-layered human oversight protocols, adaptive ethical decision-making Artificial Intelligence-based algorithms, and robust cybersecurity frameworks is proposed. Moreover, this research emphasizes the important role of meaningful human control as a fundamental safety mechanism, exploring methods to maintain effective human oversight without compromising the operational advantages of autonomy. The findings reveal the importance of a proactive, risk-based approach to AWS safety, highlighting the need for international collaboration in establishing standardized safety benchmarks and certification processes. This research contributes with valuable insights to the ongoing discourses on responsible innovation in military technology, offering evidence-based recommendations for policymakers, engineers, and ethicists working to ensure the safe and ethical development of AWS as trustworthy systems.

Keywords: Safety, Trustworthy AI, Autonomous weapons systems, Trustworthy AWS, Artificial Intelligence, Military operations, Ethics in war

1. Introduction

Autonomous Weapon Systems (AWS) have emerged as transformative technology in modern military operations, being supported by ongoing advancements in the fields of Artificial Intelligence (AI) and robotics. By enabling enhanced precision, operational efficiency, and strategic impact, these systems offer notable advantages, such as minimizing risks to human personnel, accelerating decision-making processes, and potentially ensuring greater adherence to Rules of Engagement (RoEs). Furthermore, the ability of these systems to operate in complex and dangerous settings without endangering human personnel underscores their strategic importance in modern defense strategies (Santhi et al., 2024; Maathuis and Chockalingam, 2023). The efficiency gains and technological advancements associated with AWS make them a cornerstone of future military innovation (Patil, Vidhale and Titarmare, 2024). Nevertheless, both the development and deployment of AWS is accompanied by significant ethical, safety, and legal challenges. One of the most significant concerns is the potential loss of accountability in autonomous decision-making processes (Rantanen, 2024; Maathuis, 2024), who notes the difficulty in attributing responsibility for unintended consequences, such as civilian casualties or violations of international humanitarian law. Additionally, the susceptibility of AWS to cyberattacks and adversarial manipulation poses serious operational risks, potentially leading to catastrophic outcomes if systems are compromised (Menon et al., 2024). Moreover, the delegation of lethal decision-making to machines raises ethical dilemmas about the dehumanization of warfare and the erosion of moral and societal norms (Wagner, 2014).

Central to this debate is the concept of trustworthiness, meaning building and deploying safe, responsible, and reliable systems (Li et al., 2023). In the context of AWS, these systems do not only need to demonstrate high reliability and performance, but also align with ethical principles, societal expectations, and relevant legal frameworks. Given the dual-use nature of AWS, the risks extend beyond the battlefield, affecting global stability, ethical norms, and public perception. As Horowitz (2021) notes, the development and deployment of AWS needs to be guided by stringent safety benchmarks and international collaboration to prevent accidental escalations and misuse. To these are added cyber security frameworks and solutions that need to safeguard AWS against both technical failures and intentional attacks (Huang and Lu, 2024). This directly points out the

necessity of understanding, assessing, and addressing the risks that AWS potentially have. The risks include system malfunctions, adversarial manipulation, and flawed decision-making frameworks (Riesen, 2022; Blanchard et al., 2024). These aspects show the complex and uncertain nature AWS have which reveals the interplay between technological capabilities and the need for robust regulatory and ethical frameworks. At the same time, this highlights the necessity of having a clear map of existing risks together with corresponding control measures that are applicable to them for assuring the safety of trustworthy AWS. This represents the research goal that this article aims to tackle. A systematic literature review is conducted following the PRISMA methodological approach (Denyer and Tranfield, 2009; Page et al., 2021), with the intention to formulate answers to the following research questions:

Research Question 1: What are the potential risks of developing and deploying AWS in military operations?

Research Question 2: What control measures could be considered in order to deal with the identified risks occurring when developing and deploying AWS in military operations?

The outline of this article is structured as follows. In Section 2, relevant research studies conducted in this domain are discussed. In Section 3, the methodological approach adopted and followed in this research is explained. In Section 4, the identified risks are presented in relation to technical and socio-technical dimensions relevant in this domain. In Section 5, corresponding control measures for minimizing and/or avoiding the identified risks are discussed. In the last section, concluding remarks and future research perspectives are presented.

2. Related Research

The safety and trustworthiness of AWS is a topic tackled by various studies for capturing and proposing different defence, technical, ethical, and regulatory frameworks and systems that allow understanding, dealing, and mitigating their corresponding design, development, and deployment risks through adequate control measures.

Felder (2021) emphasizes the significance of rigorous safety protocols in AWS development, proposing enhanced testing environments for evaluating dependability under combat conditions. The author stresses the fact that robust system design is critical for preventing failures that could lead to unintended consequences in military operations. Spayne et al. (2024) expand on this, suggesting that AWS should operate with a dual focus on being "safe to deploy" and ensuring "ongoing operational safety". The study introduces a regulatory approach that integrates human oversight with adaptive technologies, ensuring compliance with military safety standards. At the same time, Alquwayzani and Albuali (2024) argue for using a "zero-trust architecture" as a security model that assumes all entities are potentially compromised unless verified. This approach is aligned with the method proposed by Dharani and Kumari (2024) for federated learning systems in order to secure model training and as such safeguarding AWS from potential cyber threats. Aligned with these goals, Brooke-Holland (2023) explores the dual challenges of interoperability and cybersecurity in AWS, calling for standardized international protocols to reduce vulnerabilities and enhance operational resilience. Similarly, Huang and Lu (2024) highlight the challenges posed by cyber-physical-social system interactions, stressing the necessity of robust cybersecurity frameworks for trust enhancement. These studies collectively underline the need for a systemic approach to safeguarding AWS against intentional misuse and technical failures. Moreover, Cools and Maathuis (2024) focus on cybersecurity measures to prevent system exploitation, particularly in multi-domain operation by proposing a layered security architecture to shield AWS from both external attacks and internal failures.

The operational risk of AWS and their potential impact on strategic stability are central concerns in military discourse. Scharre (2016) analyses how autonomous systems could inadvertently escalate conflicts due to unpredictable behaviour or misinterpretation of combat scenarios. The study advocates for international agreements to mitigate such risks through standardized protocols and testing. Furthermore, Horowitz (2021) evaluates AWS's impact on deterrence and stability, arguing that system reliability is a key factor in avoiding accidental engagements which requires established rules and principles for proper target engagement.

At the same time, Kwik (2024) elaborates on adversarial countermeasures, focusing on the risks of over-reliance on autonomous decision-making in combat scenarios. Complementing these studies, ethical dilemmas surrounding the accountability and trustworthiness of AWS are directly relevant to fostering public and stakeholder confidence. To this end, Bayan et al. (2024) discuss the moral challenges inherent in AI-powered military systems, emphasizing the importance of trust in the system's ethical decision-making mechanisms. Hasan and Islam (2024) further emphasize the role of public perception and trust in determining the

acceptability of autonomous systems, emphasizing that national security considerations cannot overshadow ethical accountability. Aligned with these narratives, Harwood (2024) presents a comprehensive view of ethical frameworks designed to maintain trust in military AI, advocating for systems that allow meaningful human control during critical operations. In a similar manner, Schoenherr (2025) explores how human-machine interaction models can be tailored to promote ethical behaviour in AWS, emphasizing the critical roles that transparency and reliability have in military decision-making processes.

As this extensive literature review shows, research into the safety and trustworthiness of AWS spans diverse domains, from ethical considerations to advanced cybersecurity solutions. The integration of meaningful human control, robust governance frameworks, and cutting-edge technological innovations are seen as a consensus across studies. Nevertheless, reflecting on concrete risks and associated control measures for mitigating or avoiding these risks for AWS in a systematic way, is lacking.

3. Research Methodology

To conduct a systematic analysis on associated risks and their corresponding control measures for developing and deploying AWS in a trustworthy way, in this research the PRISMA methodological approach is adopted and followed (Denyer and Tranfield, 2009; Page et al., 2021). First, clear objectives are formulated to specify relevant risk factors and corresponding control measures for AWS in military contexts. Next, an extensive search of IEEE Xplore, ACM, SAGE Journals, and Wiley is conducted using keyword combinations such as “AWS,” “Autonomous Weapon Systems,” “trustworthy,” “trustworthiness,” “military,” and “defence.” This initial search yields a total of 10,354 studies. After removing duplicate records, a rigorous inclusion and exclusion process is carried out, ensuring that retained research studies address both AWS and trust dimensions, are written in English, and were published between 01 January 2023 and 31 December 2023. The final set of articles is then subjected to an in-depth analysis aimed at identifying patterns, evaluating control measures, and determining how risks are mitigated in various military scenarios. Accordingly, the results of this synthesis are reported in accordance with the review protocol, as illustrated in Figure 1 below:

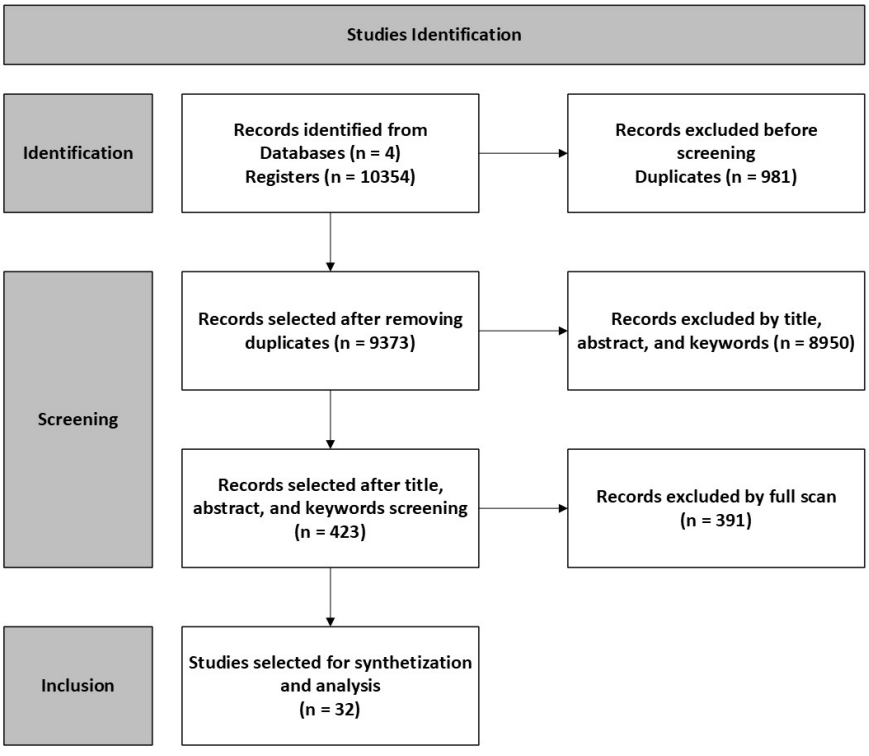

Figure 1: Research Methodology applied

4. Risks

The safety of AWS is critical for their trustworthiness, reliability, and acceptance. In this research, the identified risks as categorized as technical and socio-technical, each presenting unique challenges for designing, deploying, and maintaining these systems.

The technical risks point out existing issues in system design, operational reliability, and security vulnerabilities. These risks are interconnected with each other, at times one amplifying the other(s), thereby requiring a holistic approach to mitigation.

As Munir et al., (2021) outline, sensors, IoT devices, and UAVs (Unmanned Aerial Vehicles) generate substantial data volumes that exceed the computational capacities of edge devices. This creates a bottleneck in data processing, exacerbated by communication inefficiencies and power constraints, particularly in remote areas. These systemic shortcomings undermine the operational reliability of AWS and form the basis for additional risks. Building on this aspect, Teixeira et al. (2023) highlight the susceptibility of UAVs to hacking and misuse for illegal activities such as smuggling or privacy invasion. These risks are amplified by potential jamming attacks and electronic warfare, as discussed in Wang et al. (2023). Moreover, the complexity inherent to AWS systems, as argued by Liu et al. (2019), increases the likelihood of operational failures during reconnaissance tasks, further threatening system integrity.

Furthermore, the development process of AWS adds another risk layer. Pa and Ma (2022) detail how lengthy development cycles and unforeseen project challenges often lead to critical failures, particularly when systems struggle to adapt to dynamic operational demands. These challenges emphasize the importance of robust design and integration strategies, which are indispensable for ensuring AWS reliability. Moreover, in this process, both hardware and software vulnerabilities pose other risks and amplify existing ones related to communication lines. In these lines, Kurt et al. (2019) reflect on the potential use of UAVs for delivering chemical or biological weapons, exploiting their low-altitude flight profiles and small payloads. Such scenarios show that stringent control mechanisms and advanced detection systems are necessary in order to prevent misuse. Concomitantly, the complexity of the AWS software adds other potential threats and risks, as Gillespie (2021) stresses. Safety-critical software, often comprising millions of lines of code, is expensive to test comprehensively. This limitation leaves systems vulnerable to errors, particularly in non-deterministic decision-making processes. Further, this creates the realm for cyber security threats that aim to engage these systems with various cyber weapons. In particular, Wu et al. (2023) points out that modern UAVs are vulnerable to denial-of-service (DoS) attacks, GPS spoofing, and command injections, which can lead to crashes or system takeovers. To these are added other physical challenges as the disruption faced by UAVs from environmental factors like airflow disturbances and Doppler effects, significantly undermining system reliability (Xiao et al., 2021). Furthermore, these vulnerabilities expand into new domains of risk that relate to ongoing technological developments in the AI, IoT, and battery technology domains (Kuzmin and Znak, 2018). Given these developments, while UAV's capabilities are enhanced, the likelihood of exposure to malicious exploitation increases as well.

Socio-technical risks of AWS arise from the complex interactions between these systems and humans, regulatory frameworks, and ethical, legal, societal principles, norms, and values. These risks are interconnected, requiring careful integration of ethical, operational, and regulatory considerations to ensure the safe and trustworthy development and deployment of AWS.

The deployment of UAVs in densely populated urban areas poses immediate socio-technical challenges such as collisions, casualties, and environmental damage in such settings, necessitating stringent safety standards like Target Level of Safety (TLS) metrics to mitigate these risks (Khan et al., 2023).

Ethical and legal concerns present another critical dimension of socio-technical risks. In this sense, Emimi, Khaleel and Alkrash (2023) address the implications of using drones for lethal surveillance and targeted killings. Delegating life-and-death decisions to automated systems introduces significant ethical dilemmas and questions of accountability, especially when compliance with international laws is uncertain. Public perception further complicates these challenges, as civilian casualties and privacy violations undermine trust in AWS. The psychological impact of AWS operations on human operators is equally concerning. This happens, for instance, due to the fact that remote-controlled operations, particularly in combat zones, may lead to desensitization to the human costs of war. The video-game-like nature of drone operations trivializes the gravity of lethal decisions, diminishing the ethical boundaries that guide military engagement (Culver, 2014). At the same time, understanding and trust the behaviour of these systems in high-stakes conflict scenarios is difficult as without trust, the reliability and broader acceptance of AWS are severely compromised (Trusilo, 2023).

At the same time, the growing prevalence of UAVs in military and civilian contexts introduces additional layers of complexity. Koblentz (2020) identifies the unique security challenges posed by drone swarms, which can target multiple sensitive points at facilities designed primarily for terrestrial threats. These systems' ability to bypass traditional defences reveals the need for redefined security protocols. Moreover, Galdorisi and

Buettner (2016) raise concerns on the dangers associated with aerial robots independently determining and executing lethal actions without human oversight. Such autonomy raises accountability issues and creates potential for misuse, particularly when civilian safety is at stake.

Furthermore, public acceptance and regulatory frameworks are central to mitigating socio-technical risks. On this behalf, transparent communication and robust regulation frameworks and instrument are required to address ethical concerns and foster societal trust (Emimi, Khaleel and Alkrash, 2023; Kastan, 2013) since the danger of excessive autonomy in AWS, where increased independence could lead to overly risky missions that further complicate their safe and ethical use (Wagner, 2014; Hartmann and Giles, 2021). From a different angle, Sethu (2019) discusses the socio-economic implications of AWS deployment, including job displacement, which necessitates legislative reforms to protect affected workers and address broader societal impacts. In humanitarian contexts, Pham et al. (2022) discuss the potential of UAV-based assistive systems in rescue operations, emphasizing their value in safely navigating hazardous terrains and aiding humans. However, the absence of adequate safeguards in such high-stakes environments could lead to catastrophic outcomes, underscoring the need for stringent safety protocols.

5. Control Measures

In the previous section, identified technical and socio-technical risks were discussed. Further, in this section, corresponding control measures for reducing or avoiding them, are tackled. Specifically, technical control measures aimed at enhancing reliability, robustness, and security of AWS systems. These measures address computational challenges, navigational safety, fault tolerance, cyber security, multi-agent system behaviour, and the integration of human oversight to ensure trustworthy operations.

In order to improve computational efficiency and data management among various stakeholders and agents, decentralized processing architectures such as blockchain and edge computing are used. These paradigms bring computational tasks closer to the edge of the network, reducing latency and enhancing UAV operational efficiency in relation to the operational task that needs to be achieved. This approach enables real-time decision-making which is critical to AWS (Munir et al., 2021). To address navigation safety, Wang et al. (2023) discuss the JA-APF method, which dynamically plans paths to avoid GPS jamming zones, ensuring reliable navigation. This method balances the trade-offs of increased costs in sailing distance and time. Similarly, Liu et al. (2019) introduces fitness functions with safety factors to optimize UAV flight paths, thereby reducing collision risks and improving task efficiency. At the same time, robust fault tolerance mechanisms are essential to mitigate system vulnerabilities and ensure resilience. Accurate positioning systems and robust GPS spoofing mitigation techniques, as highlighted in (Khan et al., 2023), they have the power to protect AWS operations against environmental disruptions and malicious interference.

Furthermore, advanced training and optimization techniques also enhance AWS performance. Continuous training of autonomous agents, as suggested by Cui and Xu (2023), refines flight routes and improves both safety and operational efficiency. Complementarily, Dong, Ai and Lui (2019) highlights the role of data mining in analysing flight engagement records. This approach extracts safe behavioural patterns, akin to methodologies used in autonomous driving, to inform and enhance decision-making in AWS.

At the same time, cyber security control measures need to be considered and implemented along the life cycle of AWS. Along these lines, Wu et al (2023) introduce the ContainerDrone framework, which detects and mitigates denial-of-service (DoS) attacks by switching to secure modes upon detecting security rule violations. Moreover, Papakonstantinou et al. (2019) highlights advancements in fleet survivability, achieved through fault-tolerant and domain-aware system designs. These strategies ensure operational continuity and increase mission success rates, even under adverse conditions.

Socio-technical control measures for AWS bring together technological development and human interaction with social, legal, and ethical principles, norms, and values. These measures ensure ethical use, public trust, regulatory compliance, and seamless integration into complex human environments.

From a regulation and accountability perspective, Tyugu (2013) advocates for implementing safety protocols similar to Asimov's laws for robots to regulate agent behaviours. At the same time, emergency control mechanisms, such as forced destruction or multicast control messages, act as fail-safes against emergent risks, while self-destruction protocols for agents experiencing communication loss further mitigate potential threats and prevent cascading failures. Additionally, it is important to maintain accountability and mitigate risks associate with either a high level of autonomy or full autonomy. On this point, Stecz and Kowaleczko (2021) emphasize the importance of manual control options in AWS, enabling human operators to intervene during

mission-critical situations and prevent adverse outcomes. This human-in-the-loop approach provides an essential layer of accountability and adaptability in complex operational scenarios. Furthermore, Kurt et al. (201) emphasize the roles of international organizations such as the International Telecommunication Union (ITU-R) and the International Civil Aviation Organization (ICAO) in regulating airspace, spectrum allocation, and operational safety. These global efforts are complemented by jurisdiction-specific licensing and operational guidelines established by national authorities, ensuring AWS operations align with both local and international standards. In addition, Galdorisi and Buettner (2016) emphasize the necessity for operators and decision-makers to adhere to laws of war, treaties, and safety protocols, aligning AWS operations with global ethical standards and minimizing risks to civilian safety. Moreover, Pham et al. (2022) point out the importance of collaboration in critical missions that deploy AWS, such as hostage rescues. By combining the strengths of human decision-making with the efficiency and accessibility of autonomous systems, these collaborations enhance safety and operational effectiveness.

Ensuring system reliability is another important control measure from a socio-technical standpoint. To this end, Gillespie (2021) recommends defining critical components as safety-critical, subjecting them to rigorous testing and stringent legal standards. In this case, non-critical components should also follow predefined trial protocols after updates to minimize risks from software modifications. Furthermore, Tyugu (2013) proposes implementing constraints on multi-agent systems to regulate behaviours and prevent harmful actions, ensuring alignment with societal values. Emergency protocols, such as forced destruction and multicast control messages, provide safeguards against agent deviations. At the same time, Papakonstantinou et al. (2019) stresses out the importance of designing AWS fleets to balance operational effectiveness with resource constraints, aligning operations with societal and mission-specific requirements. Finally, the integration of AWS into shared environments necessitates comprehensive safety and operational standards. In this respect, Pham et al. (2022) illustrate the value of designing collaborative systems where human operators and autonomous agents work synergistically. This ensures AWS contributions complement human capabilities while adhering to ethical and safety standards.

6. Conclusions

The development and deployment of AWS represents an important leap in the military domain, offering precision, operational efficiency, and the potential to reduce human casualties on the battlefield. Nevertheless, these advancements bring significant challenges, particularly regarding safety, ethical considerations, and socio-technical complexities. Ensuring the trustworthiness of AWS is necessary as the risks associated with system malfunctions, unintended engagements, flawed ethical decision-making, and cyber security incidents. These risks can have various implications and consequences for combatants, civilians, and international security (Maathuis, 2024b). In particular, safety serves as a foundational pillar for establishing trust, implying the integration of fail-safe mechanisms, human oversight protocols, adaptive ethical AI systems, and robust cybersecurity frameworks and solutions. In this domain, this research conducts a systematic literature review to address these critical challenges, analysing safety risks through from a double perspective: (i) of technical and (ii) of socio-technical dimensions while proposing measures to enhance AWS reliability and accountability. Given these results, future research could focus on building modelling solutions that contribute to developing standards to ensure both alignment and consistency across international deployments. Additionally, investigating innovative methods for real-time human-AI collaboration could optimize decision-making processes while maintaining the advantages of autonomy with considerations to aspects such as feedback and trust. These aspects are important to advancing AWS as safe, responsible, and trustworthy systems that are aligned with the demands of current warfare perspectives and international norms.

References

Alquwayzani, A. A., & Albuali, A. A. (2024). A Systematic Literature Review of Zero Trust Architecture for Military UAV Security Systems. *IEEE Access*.

Ashokkumar, C. R. (2019). Unmanned air vehicles for targeting tasks. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *233*(5), 1926-1934.

Bayan, F. M. (2024). The Ethics of AI: Navigating the Moral Dilemmas of Artificial Intelligence. *Arab Journal for Scientific Publishing (AJSP) ISSN*, *2663*, 5798.

Blanchard, A., Novelli, C., Floridi, L., & Taddeo, M. (2024). A Risk-Based Regulatory Approach to Autonomous Weapon Systems. *Available at SSRN*.

Brooke-Holland, L. (2023). AUKUS Pillar 2: Advanced Capabilities Programmes. *Research Briefing*.

Cools, K., & Maathuis, C. (2024, October). Trust or Bust: Ensuring Trustworthiness in Autonomous Weapon Systems. In *MILCOM 2024-2024 IEEE Military Communications Conference (MILCOM)* (pp. 182-189). IEEE.

Cui, Q., & Xu, K. (2023). A Hierarchical Framework for Multi-UAV Reconnaissance Mission Planning Problem. In *2023 IEEE 3rd International Conference on Electronic Technology, Communication and Information (ICETCI)* (pp. 1616-1623). IEEE.

Culver, K. B. (2014). From battlefield to newsroom: Ethical implications of drone technology in journalism. *Journal of mass media ethics*, *29*(1), 52-64.

Denyer, D. and Tranfield, D. (2009). Producing a systematic review.

Dharani, D., & Anitha Kumari, K. (2024). A smart surveillance system utilizing modified federated machine learning: Gossip-verifiable and quantum-safe approach. *Concurrency and Computation: Practice and Experience*, *36*(24), e8238.

Dong, Y., Ai, J., & Liu, J. (2019). Guidance and control for own aircraft in the autonomous air combat: A historical review and future prospects. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, *233*(16), 5943-5991.

Emimi, M., Khaleel, M., & Alkrash, A. (2023). The current opportunities and challenges in drone technology. *Int. J. Electr. Eng. and Sustain.*, 74-89.

Felder, J. A. (2021). Safety Engineering of Weaponized Autonomous Systems. Doctoral dissertation, Monterey, CA; Naval Postgraduate School).

Fish, A., & Richardson, M. (2022). Drone power: conservation, humanitarianism, policing and war. *Theory, Culture & Society*, *39*(3), 3-26.

Galdorisi, G., & Buettner, R. (2016). Bridging multiple autonomous vehicle disciplines: Ensuring autonomous systems promote peace and stability on the world's oceans. In *OCEANS 2016 MTS/IEEE Monterey* (pp. 1-13). IEEE.

Gillespie, T. (2021). Good practice for the development of autonomous weapons: Ensuring the art of the acceptable, not the art of the possible. *The RUSI Journal*, *165*(5-6), 58-67.

Hartmann, K., & Giles, K. (2016). UAV exploitation: A new domain for cyber power. In *2016 8th international conference on cyber conflict (CyCon)* (pp. 205-221). IEEE.

Harwood, S. (2024). A cybersystemic view of autonomous weapon systems (AWS). *Technological Forecasting and Social Change*, *205*, 123514.

Hasan, M. M. U., & Islam, M. S. (2024). The Role of Artificial Intelligence in Military Systems: Impacts on National Security and Citizen Perception.

Horowitz, M. C. (2021). When speed kills: Lethal autonomous weapon systems, deterrence and stability. In *Emerging technologies and international stability* (pp. 144-168). Routledge.

Huang, Y., & Lu, X. (2024). Security, governance, and challenges of the new generation of cyber-physical-social systems. Frontiers in Physics, 12, 1464919.

Jiang, D., & Peng, Q. Y. (2023, August). Analysis of Structural Characteristics of Water Traffic System Based on OPDAR Model. In *2023 7th International Conference on Transportation Information and Safety (ICTIS)* (pp. 538-544). IEEE.

Kastan, B. (2013). Autonomous Weapons Systems: A Coming Legal" Singularity"?. *U. Ill. JL Tech. & Pol'y*, 45.

Khan, A., Campos, J. R., Ivaki, N., & Madeira, H. (2023). A Machine Learning driven Fault Tolerance Mechanism for UAVs' Flight Controller. In *2023 IEEE 28th Pacific Rim International Symposium on Dependable Computing (PRDC)* (pp. 217-227). IEEE.

Kuzmin, A., & Znak, E. (2018). Blockchain-base structures for a secure and operate network of semi-autonomous unmanned aerial vehicles. In *2018 IEEE International conference on service operations and logistics, and informatics (SOLI)* (pp. 32-37). IEEE.

Kwik, J. (2024). Adversarials: Anti-AI Countermeasures. In *Lawfully Using Autonomous Weapon Technologies* (pp. 129-155). The Hague: TMC Asser Press.

Koblentz, G. D. (2020). Emerging technologies and the future of CBRN terrorism. *The Washington Quarterly*, *43*(2), 177-196.

Kurt, G. K., Khoshkholgh, M. G., Alfattani, S., Ibrahim, A., Darwish, T. S., Alam, M. S., ... & Yongacoglu, A. (2021). A vision and framework for the high altitude platform station (HAPS) networks of the future. *IEEE Communications Surveys & Tutorials*, *23*(2), 729-779.

Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., ... & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, *55*(9), 1-46.

Liu, C., Xie, W., Zhang, P., Guo, Q., & Ding, D. (2019, June). Multi-uavs cooperative coverage reconnaissance with neural network and genetic algorithm. In *Proceedings of the 2019 3rd High Performance Computing and Cluster Technologies Conference* (pp. 81-86).

Maathuis, C., & Chockalingam, S. (2023). Modelling the influential factors embedded in the proportionality assessment in military operations. In *International Conference on Cyber Warfare and Security* (Vol. 18, No. 1, pp. 218-226).

Maathuis, C. (2024). Trustworthy Human-Autonomy Teaming for Proportionality Assessment in Military Operations. In *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)* (pp. 1-8). IEEE.

Maathuis, C. (2024b). Towards Trustworthy AI-based Military Cyber Operations. In *International Conference on Cyber Warfare and Security* (Vol. 19, No. 1, pp. 129-136).

Menon, S., Todariya, S., & Agerwala, T. (2024). Fundamental Reflections on Minds and Machines. In *AI, Consciousness and The New Humanism: Fundamental Reflections on Minds and Machines* (pp. 1-9). Singapore: Springer Nature Singapore.

Munir, A., Kwon, J., Lee, J. H., Kong, J., Blasch, E., Aved, A. J., & Muhammad, K. (2021). FogSurv: A fog-assisted architecture for urban surveillance using artificial intelligence and data fusion. *IEEE Access*, *9*, 111938-111959.

Pan, Q., & Ma, Z. (2022). Research and development of mosaic warfare.

Papakonstantinou, N., Bashir, A. Z., O'Halloran, B., & Van Bossuyt, D. L. (2019). Early assessment of drone fleet defence in depth capabilities for mission success. In *2019 Annual Reliability and Maintainability Symposium (RAMS)* (pp. 1-7). IEEE.

Patil, A. T., Vidhale, B., & Titarmare, A. (2024). Strategic innovations in defense systems: a comprehensive analysis of emerging technologies and future trends. In *2024 International Conference on Intelligent Systems for Cybersecurity (ISCS)* (pp. 1-7). IEEE.

Page, M.J. *et al.* (2021) 'Declaración PRISMA 2020: una gu\'\ia actualizada para la publicación de revisiones sistemáticas', *Revista española de cardiolog\'\ia*, 74(9), pp. 790–799.

Pham, D., Menon, V., Tenhundfeld, N., Weger, K., Mesmer, B., Gholston, S., & Davis, T. (2022). A Case Study of Human-AI Interactions Using Transparent AI-Driven Autonomous Systems for Improved Human-AI Trust Factors. In *2022 IEEE 3rd International Conference on Human-Machine Systems (ICHMS)* (pp. 1-6). IEEE.

Rantanen, A. (2024). Are Individuals in an Armed Conflict Developing into Zeroes and Ones?: A Study of Autonomous Weapon Systems and International Humanitarian Law.

Riesen, E. (2022). The moral case for the development and use of autonomous weapon systems. *Journal of Military Ethics*, *21*(2), 132-150.

Santhi, K., Shri, M. L., Joshi, S., & Sharma, G. (2024, February). AI in Defence and Ethical Concerns. In *2024 Second International Conference on Emerging Trends in Information Technology and Engineering (ICETITE)* (pp. 1-7). IEEE.

Scharre, P. (2016). *Autonomous weapons and operational risk*.

Schoenherr, J. R. (2025). Meaningful Human Control of Autonomous Weapons Systems: Translating Functional Affordances to Inform Ethical Assessment and Design. In *Cyber Security in the Age of Artificial Intelligence and Autonomous Weapons* (pp. 173-199). CRC Press.

Sethu, S. G. (2019). The inevitability of an international regulatory framework for artificial intelligence. In *2019 International Conference on Automation, Computational and Technology Management (ICACTM)* (pp. 367-372). IEEE.

Spayne, P., Lacey, L., Cahillane, M., & Saddington, A. (2024). Operating itself safely: merging the concepts of 'safe to operate'and 'operate safely'for lethal autonomous weapons systems containing artificial intelligence. *Defence Studies*, 1-35.

Stecz, W., & Kowaleczko, P. (2021). Designing Operational Safety Procedures for UAV According to NATO Architecture Framework. In *ICSOFT* (pp. 135-142).

Teixeira, K., Miguel, G., Silva, H. S., & Madeiro, F. (2023). A survey on applications of unmanned aerial vehicles using machine learning. *IEEE Access*.

Thompson, F., & Guihen, D. (2019). Review of mission planning for autonomous marine vehicle fleets. *Journal of Field Robotics*, *36*(2), 333-354.

Trusilo, D. (2023). Autonomous AI systems in conflict: Emergent behavior and its impact on predictability and reliability. *Journal of Military Ethics*, *22*(1), 2-17.

Tyugu, E. (2013). Situation awareness and control errors of cyber weapons. In *2013 IEEE International Multi-Disciplinary Conference on Cognitive Methods in Situation Awareness and Decision Support (CogSIMA)* (pp. 143-148). IEEE.

Wagner, M. (2014). The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapon systems. *Vand. J. Transnat'l L.*, *47*, 1371.

Wang, J., Xiao, Y., Li, T., & Chen, C. P. (2023). A jamming aware artificial potential field method to counter GPS jamming for unmanned surface ship path planning. *IEEE Systems Journal*, *17*(3), 4555-4566.

Wu, S., Li, Y., Wang, Z., Tan, Z., & Pan, Q. (2023). A highly interpretable framework for generic low-cost UAV attack detection. *IEEE Sensors Journal*, *23*(7), 7288-7300.

Xiao, Z., Zhu, L., Liu, Y., Yi, P., Zhang, R., Xia, X. G., & Schober, R. (2021). A survey on millimeter-wave beamforming enabled UAV communications and networking. *IEEE Communications Surveys & Tutorials*, *24*(1), 557-610.