Faculty of Humanities and Social Sciences

Master in International Security Management

Final thesis

To what extent does the integration of artificial intelligence into autonomous weapons challenges military ethics and international security?

Clara Núñez Vázquez

2025-2026

In this thesis I present the integration of artificial intelligence into autonomous weapons and how these clashes with military ethics and international security. Literature is reviewed with the objective of understanding the issues that scholars have raised. I give a guideline of how methodology should be applied in order to explore this topic to get the best conclusions touching upon different relevant aspects.

Key words: Artificial Intelligence, Autonomous Weapons Systems, Military, NATO, ethics.

INDEX

INTRODUCTION

One of the most important current discussions in military ethics and global security is the incorporation of artificial intelligence into autonomous weapon systems. Concerns regarding the implications of decreasing human involvement in life-or-death decisions have been voiced by academics and policymakers as AI-enabled systems grow more capable of supporting or performing tasks linked to targeting, surveillance, navigation, and engagement. Researchers draw attention to a number of interrelated problems, such as potential deterrence stability and escalation concerns, accountability gaps in circumstances of unlawful harm, conflicts with the principles of difference and proportionality, and the potential degradation of effective human control. As NATO and its member states gradually incorporate AI and autonomy into defense planning and operational procedures, it is imperative to determine whether these technological advancements can be balanced with ethical and legal commitments.

The main research question this thesis attempts to answer is "*to what extent does the integration of artificial intelligence into autonomous weapons challenges military ethics and international security?*". Three sub-questions are taken into consideration in order to direct the analysis: how AI-driven autonomy impacts well-established military ethics principles like accountability and meaningful human control; what effect AI-enabled systems may have on deterrence stability and the risk of escalation; and how NATO and EU regulatory frameworks handle the ethical issues raised by autonomous weapons. The study is based on the fact that, despite the large body of material on autonomous weapons, a large portion of it is theoretical and does not adequately address the practical application of ethical standards within alliances like NATO.

This thesis uses a qualitative approach based on document analysis and case study evaluation to close this gap. This approach is suitable since the topic is normative, regulatory, and interpretive, necessitating a review of institutional frameworks, policy texts, and ethical considerations rather than depending solely on statistical data. Four evaluation criteria are applied in the analysis, which builds on the literature review: the preservation of meaningful human control over lethal decisions; accountability and clarity of responsibility; adherence to international humanitarian law and human rights standards; and the implications for deterrence stability and escalation risk.

The two practical case studies evaluated against these criteria (together with NATO and EU policy frameworks) are the deployment of the AI-enabled Merops counter-drone system by NATO and the Lithuanian GA-10FPV-AI (with related platforms like Hornet XR). While the GA-10FPV-AI is an offensive, strike-capable FPV drone created by a lesser Ally and given to Ukraine, Merops is an example of a defensive counter-UAS capability used by larger Allies along NATO's eastern flank. Together, they cover a variety of complementary but distinct uses of AI-enabled autonomy, including as precise strikes, reconnaissance, and air defense interception.

Different institutional and strategic circumstances are also reflected in the cases. The integration of Merops into NATO-affiliated air defense systems in Poland, Romania, and Denmark demonstrates the application of Alliance-level principles in a global context. In contrast, the GA-10FPV-AI program demonstrates how a smaller Ally both utilizes and sells AI-enabled drones, raising concerns about export regulations, traceability, and the spread of NATO and EU standards outside of official Alliance activities. When taken as a whole, these cases offer a strong foundation for assessing whether current governance frameworks maintain the compatibility of AI integration into autonomous weapons with international security, international humanitarian law, and military ethics.

LITERATURE REVIEW

This literature review is going to be organized in certain subtopics addressing relevant literature related to this topic of research. First, ethical and moral aspects are going to be considered. To follow up, legal frameworks and challenges will be addressed through literature, and lastly, the security implication of this issue will be analysed.

2.1 Ethical and Moral Implications

Advancements in autonomous weapons on the military field have raised ethical and legal concerns among scholars. AI-enabled autonomous weapon systems (AWS) have led to a loss of meaningful human control (MHC). Some of the main issues that have been raised are the violation of human dignity of possible victims of autonomous weapons systems, the consideration for laws of war, responsibility ascription matters, and overall, a rise of global instability. The main question in this discussion is whether autonomy can ever coexist with moral precepts like proportionality and differentiation, which have historically required human judgment. Daniele Amoroso's research, for instance, emphasizes that early discussions position meaningful human control as not only desirable but essential to ethical and legal legitimacy, highlighting that even legal approval of AWS depends on keeping human oversight throughout targeting and engagement. (Amoroso & Tamburrini, 2020)

This matter is also explored by scholars arguing that there is not only an ethical issue, but also a practical one. They argue how it is difficult to hold actors accountable for disproportionate attacks with the use of AI since Artificial Intelligence cannot bear any moral or legal liability. This would lead to responsibility gaps. This is addressed by academics such as Mengkorn Pum, also touching upon the concern for an arms race due to the lack of regulation on the use of AWS. He implies how there must be a human supervision of this systems in order to avoid mistakes and to ensure the compliance of international regulations during conflict. (Pum, 2024)

Why issues of human control and dignity are so important in today's discussions is further explained by a Kantian viewpoint on autonomous weapon systems. According to Kant, moral autonomy is the foundation of human dignity. Since individuals are competent to rationally legislate themselves, they must always be viewed as ends in and

of themselves rather than just as means. In the context of AWS, a number of authors contend that giving non-moral agents the authority to make life-or-death decisions compromises this condition because targets are no longer acknowledged as bearers of intrinsic worth but rather are reduced to objects processed by an algorithm. Since no machine can have the kind of moral agency that Kantian dignity is founded upon, eliminating or drastically reducing human judgment at the point of lethal decision-making is not only a technical or legal issue, but also a violation of the fundamental moral architecture that underpins humanitarian law (Ulgen, 2017).

Another normative framework for assessing AI-enabled autonomy in combat is provided by just war theory, especially in its modern articulation by Michael Walzer. Conventional jus in bello concepts like difference, proportionality, and accountability require agents to be able to identify morally significant aspects of a situation, use sound judgment, and be held responsible for transgressions. By introducing non-human decision-making into targeting and escalation procedures, autonomous systems complicate this framework and cast question on the continued reliability of fundamental just war principles in intricate, dynamic contexts. Given that NATO doctrines aim to incorporate these ideas into the rules of engagement, the increasing use of AWS and AI-assisted targeting points to a conflict between the desire to fight "just" wars and the realities of implementing just war principles in an era where deadly decisions are increasingly made by machines rather than combatants (Lazar, 2017).

By concentrating on how NATO's desire for AI-enabled autonomy intersects with fundamental moral precepts like just war theory and international humanitarian law, researchers broaden the ethical discussion. Sharkey cautions that as decision-making becomes more detached from human judgment, increased automation in NATO operations runs the risk of undermining the idea of moral agency. She contends that the increasing dependence on machine-driven targeting systems creates a moral and psychological barrier between operators and the results of their actions, even if humans are still "on the loop." According to Sharkey, this disengagement could jeopardize the proportionality principle and the moral obligation to stop needless injury, especially when AI technologies are used in delicate settings with unpredictable civilian presence (Sharkey, 2019).

Collectively, these writers illustrate that the diminishing of human supervision undermines the ethical basis of combat. Amoroso underscores the importance of maintaining significant control as a moral foundation, whereas Sharkey (2019) and Pum (2024) point out that automation can dull operators' sensitivity and blur accountability. This merging of ethical and practical risks emphasizes that ensuring human accountability is both a legal and a moral necessity in military operations.

Nonetheless, although there is consensus on the significance of human supervision, limited research outlines what "meaningful control" should practically entail or how organizations like NATO can consistently implement it among its members. This research thus investigates how NATO's ethical guidelines tackle this ambiguity and if they can effectively transform moral values into operational criteria.

2.2 Legal and Regulatory Challenges

Autonomous weapons systems (AWS) provide significant ethical and governance issues that go beyond technical performance, according to existing research. Delegating lethal decisions to computers can undermine accountability frameworks in armed conflict and decrease the moral responsibility often held by human decision-makers, according to research on AWS ethics. According to a comparative analysis of AWS definitions, variations in how states and organizations define autonomy have an impact on how ethical and legal issues are perceived and handled, making it challenging to reach a consensus and resulting in normative gaps in deployment and regulation. This ambiguity is especially pertinent to NATO, as member states differ in their ethical control doctrines and interpretations, making Alliance-wide ethical consistency on autonomous systems more difficult to achieve (Taddeo, 2022).

The necessity for frameworks that ground technological integration in fundamental ethical concepts like accountability, proportionality, and transparency is further highlighted by academic research on the ethics of military AI. When AI is incorporated into systems that support or make decisions in life-critical situations, ethical principles like fairness, accountability, transparency, and explainability (FATE) are crucial, according to a recent ethical study of AI use in military visual reconnaissance. In addition to adhering to legal standards like International Humanitarian Law (IHL), the

study contends that military applications of AI should incorporate moral standards specific to how autonomous systems perceive and behave in complicated contexts. This supports arguments that ethical governance, even in alliances like NATO, needs operationalized ethical standards and human monitoring procedures to guarantee the alignment of values between AI behavior and human moral expectations. (Dr. Burkart & Jeschke, 2025).

Ultimately, human rights organizations question NATO's moral stance. It is argued that without legal restrictions on fully autonomous weapons, NATO's commitment to "responsible use" is inadequate. They contend that since no machine is capable of moral judgment or empathy, giving fatal force to machines breaches fundamental human rights, especially the right to life and due process. These organizations support a legally binding international agreement that would completely prohibit lethal autonomous weaponry (Stauffer, 2025).

The literature in security studies examines how AWS and AI integration affects international security, conflict escalation, and strategic stability in addition to ethics and law. Because decision-making times are getting shorter and uncertainty is increasing, the quick development and implementation of AI-driven systems like swarms and drones may change military strategy and open up new avenues for conflict escalation. According to some theories, autonomous weapons could lower the threshold for war by lowering the dangers to human soldiers and speeding up lethal action, which could unintentionally increase the frequency of conflicts or make them more difficult to govern. (Burrows, 2026).

Research on AI-powered drone swarms shows how these technologies have the potential to change combat dynamics by facilitating adaptive reactions and decentralized cooperation in the absence of centralized human orders. According to research, swarms' capacity for independent decision-making raises significant concerns regarding how commanders can maintain responsibility and control as well as how adversaries can react to such quickly changing threats (Malyasov, 2025).

Taddeo and Blanchard (2022) highlight that varying definitions of AWS among NATO countries have led to regulatory fragmentation, complicating the establishment of ethical consistency across the alliance. These results are crucial to this study’s emphasis

on NATO as they highlight the difficulty of establishing a cohesive legal and ethical framework among member nations. The study seeks to assess whether the alliance's existing governance can adequately tackle the legal uncertainty regarding autonomous weapons.

2.3 Security Implications

To follow up on this, human rights advocates have expanded the literature to contend that AWS poses a threat to fundamental human rights principles in addition to international security and battlefield ethics. According to reports by human rights organizations, the delegation of life-and-death decisions by autonomous systems violates fundamental rights like the right to life, non-discrimination, and due process since computers lack the judgment and contextual awareness required to protect these rights. One significant research highlights the inherent challenges autonomous systems encounter in comprehending complex social situations and describes how autonomous targeting systems may breach international human rights commitments both in times of war and peace (Stauffer, 2025).

Recent research has increasingly concentrated on how organizations like NATO handle the moral and normative issues raised by AI-driven autonomy in combat, building on these larger ethical and human rights discussions. Scholars are starting to wonder if NATO's frameworks for autonomous systems and responsible AI use sufficiently protect human oversight, responsibility, and moral constraint in military operations.

Scholars and policymakers are debating ethics and norms in response to NATO's growing dependence on AI-driven autonomy in military operations. Although the Alliance formally supports the responsible integration of AI, scholars continue to raise questions about how NATO's frameworks balance human morality and technology efficiency. AI-enabled drones pose concerns about proportionality, discrimination, and the morality of removing human operators from deadly decisions, according to the NATO Review article "Autonomous Military Drones: No Longer Science Fiction." The paper emphasizes the risk that automation could make commanders less morally restrained in combat by desensitizing them to the use of force (NATO, 2017).

The literature collectively acknowledges that incorporating AI into military operations has significant effects on global stability. Nevertheless, there is a lack of

empirical analyses regarding the impact of these technologies on security architectures based on alliances. This gap underscores the necessity of this study's case-based method, which will assess the tangible effects of AI-powered systems like the Merops and Lithuanian drones on NATO's ethical and strategic equilibrium.

Although there is significant discussion regarding the ethics and legality of autonomous weapons, studies rarely investigate how these principles are practically implemented in NATO or other alliances. Current research tends to be largely theoretical, highlighting a lack of insight into how ethical principles like accountability and meaningful human control operate in practical AI applications. This research tackles that void with a detailed examination of NATO's governance structures and case analyses.

RESEARCH METHODOLOGY

In order to address the research question "**to what extent does the integration of artificial intelligence into autonomous weapons challenges military ethics and international security?**" this study employs a qualitative technique based on document analysis and case study evaluation. Since the subject being studied (AI-driven autonomy in weapons) is considered as normative, regulatory, and interpretive, a qualitative method is most suited for this essay. Examining ethical considerations and legal frameworks is necessary for the research rather than depending solely on statistical data. A qualitative method is selected since it facilitates the analysis of policy documents and discursive materials to comprehend how ethical principles are understood and implemented in practice

3.1 Policy analysis

The ethical and security issues raised by AI-enabled autonomous weapon systems are examined in this section along with the extent to which current NATO and EU policy frameworks operationalize important concepts like meaningful human control, accountability, and adherence to international humanitarian law (IHL). Building on the literature review, the analysis uses four evaluation criteria that are drawn from official doctrine and scholarly discussion: (1) preservation of meaningful human control over lethal decisions; (2) accountability and clarity of responsibility; (3) compliance with international humanitarian law and human rights standards; and (4) implications for deterrence stability and escalation risk. The two empirical case studies (Merops and Lithuanian drones), which serve as real-world tests of how far these policy commitments can be realized in practice, are linked to the main strategies, implementation plans, and political declarations of NATO and the EU after being evaluated against these criteria.

3.2 Operationalization of analytical criteria

The four analytical criteria are considered variables with specific observable indicators for each of the two case studies (Merops and the GA-10FPV-AI drones). This enables the qualitative analysis to go beyond broad normative claims and evaluate, in an organized manner, the extent to which the governance structures of the EU and NATO are actually put into reality.

- *The degree to which human operators maintain effective control* over the use of deadly force during the targeting and engagement cycle is measured by the meaningful human control criterion. The ability of human operators to approve, alter, or terminate engagements; the degree of human involvement in target selection and engagement rules; the ability of operators and commanders to comprehend and override system recommendations; and the degree of autonomy exercised by the system during mission execution (e.g., human-in-the-loop, human-on-the-loop, or higher levels of automation) are some indicators.

- *Accountability and clarity of responsibility*: this criterion evaluates how blame is assigned in the event of malfunction or unlawful injury. The following are examples of indicators: (a) the presence of clearly defined chains of command and documented rules of engagement; (b) the distribution of accountability among developers, national authorities, and NATO structures in official texts; (c) the existence of mechanisms (like logs or audit trails) to track decisions made by human operators and AI modules; and (d) references to sanctions or corrective actions in the event of violations.

- *Compliance with international humanitarian law and human rights standards*: this criterion assesses whether the systems' implementation and design clearly uphold pertinent human rights commitments as well as fundamental IHL principles including distinction, proportionality, and precaution. The following are indicators: (a) whether the systems are subject to Article 36 weapons reviews or comparable national legal assessments; (b) technical features (such as sensor suites, identification protocols, or fail-safe modes) intended to improve discrimination and reduce collateral damage; (c) operational guidelines on the use of the systems in complex environments like urban areas or mixed civilian-military airspace; and (d) references to human rights or civilian-protection concerns in official rhetoric surrounding their deployment.

- *Implications for escalation risk and deterrence stability:* this criterion looks at how the implementation of AI-enabled technologies impacts both the danger of unintentional escalation and crisis stability. The role of the systems in national and NATO deterrence concepts (e.g., deterrence by denial versus deterrence by punishment); (b) the degree to which increased speed, automation, and persistence

may shorten decision-making timelines; (c) the possibility of misperception or unintentional engagements in high-intensity or sensitive border environments; and (d) whether official documents or expert assessments mention safeguards or confidence-building measures to mitigate escalation risks.

In order to give a comparative evaluation of how AI-enabled systems perform in relation to the four criteria, these indicators are applied systematically to the Merops and GA 10FPV AI instances in the empirical chapters and revisited in the discussion section.

The thesis distinguishes between human-in-the-loop, human-on-the-loop, and human-out-of-the-loop systems using a standard taxonomy of autonomy commonly employed in the context of the Convention on Certain Conventional Weapons (CCW) to prevent conceptual ambiguity. Human-out-of-the-loop systems choose and use force without significant real-time human supervision; human-on-the-loop systems permit autonomous operation but maintain a human supervisor who can monitor and intervene; and human-in-the-loop systems require a human decision to start every engagement. In order to more clearly evaluate how varying levels of autonomy impact meaningful human control, responsibility, legal compliance, and deterrence dynamics, the two case studies are categorized along this continuum in the empirical chapters.

3.3 NATO's governance of AI and autonomy

NATO has so far chosen a non-binding, principles-based strategy over a legally binding pact on autonomous weapons. The 2024 updated Artificial Intelligence Strategy, which outlines six Principles of Responsible Use (PRUs) for all AI-enabled systems (lawfulness, responsibility, explainability, dependability, governability, and bias mitigation) is the cornerstone of this strategy (NATO, 2024). These PRUs are designed to guarantee that AI-enabled capabilities, such as deadly autonomous weapon systems, function in accordance with shared Alliance principles and international law. The 2022 Autonomy Implementation Plan, which outlines how member states should develop, test, and implement autonomous and semi-autonomous systems in accordance with IHL, common standards, and interoperability requirements, simultaneously transforms the overall strategic direction into more practical terms (NATO, 2022).

NATO has also started to establish institutional oversight procedures. The establishment of the Data and AI Review Board (DARB), which is meant to oversee the certification and supervision of AI use in defense with a focus on system traceability, documenting of design decisions, and human monitoring over crucial functions, is especially pertinent. The PRUs, the Autonomy Implementation Plan, and DARB collectively indicate a clear normative intention: AI and autonomy should be integrated into NATO capabilities in a way that protects human judgment, guarantees legal compliance, and permits post-hoc evaluation in the event that something goes wrong (Arrows, 2025).

However, significant gaps and uncertainties appear when these instruments are assessed using the four criteria. First, even though NATO frequently emphasizes the significance of meaningful human control, it does not offer a clear operational description of what such control necessitates in various system categories (defensive vs. offensive, strategic vs. tactical, human-in-the-loop vs. on-the-loop). Taddeo and Blanchard's comparative research demonstrates that definitions of autonomous weapon systems already differ greatly amongst states, including NATO nations; this definitional fragmentation makes it more difficult for the Alliance to guarantee uniform ethical standards in cooperative operations. In actuality, this means that both Allies can claim to adhere to NATO values even while what one Ally views as enough human oversight may be deemed inadequate by another.

Second, although the PRUs discuss accountability and governability, NATO's documents essentially delegate actual accountability to individual member states. The Alliance retains sovereign authority over deployment choices, weapons reviews, and rules of engagement; it lacks a legally enforceable enforcement mechanism and the authority to discipline infractions. This presents difficult issues regarding accountability in joint missions: is the deploying state, the developer state, or the NATO command structure ultimately responsible if an AI-enabled machine does illegal harm? According to the research on responsibility gaps in autonomous weapons, this ambiguity is not only theoretical and may jeopardize public confidence as well as legal accountability.

Third, it is still unclear how NATO's tenets relate to deterrence and escalation dynamics. Although NATO recognizes in policy discussions that quicker decision-making and more automation may change the stability of a crisis, official documents place

more emphasis on preserving technological advantage and interoperability than on outlining specific precautions against unintentional escalation, such as in human-machine interaction at high operational tempo. The Merops instance, where autonomous counter-drone defense along the eastern flank is justified in terms of cost-effectiveness and robustness under electronic warfare conditions, will highlight this conflict between strategic imperatives and ethical constraint.

When considered collectively, NATO's framework might be characterized as institutionally light yet normatively ambitious. Strong rhetorical commitments to human control, legality, and responsibility are embedded, but member states are given considerable latitude in interpreting and carrying out these obligations. This creates the possibility of a "lowest common denominator" dynamic, in which the standards used in coalition operations may actually be shaped by the Allies who are most at ease with high degrees of autonomy and laxer interpretations of human supervision. When evaluating how Merops and the Lithuanian drones have been incorporated into NATO's overall stance, the policy analysis will revisit this subject.

3.4 EU norms and the European governance gap

Although the European Union is not a defense alliance, many NATO members operate in a normative framework that is greatly influenced by its regulatory decisions. The European Parliament has consistently demanded "meaningful human control" over lethal force and international restrictions on fully autonomous weapons, while the majority of EU and NATO governments explicitly share a commitment to IHL and the Convention on Certain Conventional Weapons (CCW) (European Union, 2021). However, the Artificial Intelligence Act, the EU's premier tool for AI governance, expressly excludes military and national security applications from its purview, giving national governments and NATO primary authority to regulate autonomous weaponry (Guedón, 2024).

As a result, the governance environment becomes complex and somewhat fractured. On the one hand, EU institutions create soft norms and political pressure that support strict human control and warn against completely autonomous deadly systems. However, there isn't a single European legal standard for AWS because there isn't a legally enforceable EU-level law for defense AI. Different national approaches are taken

by member states. France and Germany have been among the clearest advocates of a two-tier model: a prohibition on fully autonomous lethal weapons that cannot comply with international humanitarian law, combined with regulation of other autonomous systems to ensure meaningful human control. At the EU level, the common baseline remains that human control must be retained over decisions on the use of lethal force, but national positions still differ in emphasis and in their preferred regulatory instruments.

The AI Act has an indirect impact even though it does not directly control systems like Merops or the Lithuanian GA-10FPV-AI. Its criteria for risk management, transparency, and human oversight apply to dual-use AI providers that service both the civilian and defense industries. These requirements may have an impact on defense applications through corporate policies and technological standards. The operational situations in which NATO-linked autonomous systems are employed can also be impacted by EU export-control regimes and political conditionality, which can influence how European governments and businesses transfer AI-enabled technologies to allies like Ukraine.

The main argument of this thesis is that soft-law concepts and instruments are crucial to both NATO and the EU. There is no cohesive, legally binding European stance on the acceptable level of autonomy in military systems, notwithstanding rhetorical consensus, particularly in the areas of human dignity, accountability, and proportionality. In the Lithuanian scenario, when a smaller ally creates and exports a significant quantity of AI-enabled drones that are continuously enhanced in combat situations in Ukraine, this governance gap is very pertinent. When systems change quickly in theater rather than through conventional procurement channels, the lack of a clear shared standard raises concerns about how Alliance-wide ethical expectations are enforced.

3.5 Policy coherence, accountability and case-study relevance

A conflicting picture of policy coherence and accountability appears when NATO and EU policy tools are taken into account collectively. On paper, both frameworks state that operations must adhere to IHL, that lethal force must stay under human control, and that new technologies should not undermine fundamental rights or strategic stability. An emerging governance architecture is indicated by NATO's PRUs, the Autonomy Implementation Plan, and the establishment of DARB, while national initiatives in

nations like France and Germany and resolutions of the EU Parliament advocate for stringent restrictions on fully autonomous systems and unambiguous human authority over targeting decisions.

However, the methods for converting these ideas into legally binding restrictions are still in their infancy. Current policy tools, such as Article 36 weapons reviews, national export controls, and Alliance-level oversight bodies, offer significant safeguards, but they were not created with highly adaptive AI-enabled systems in mind. As a result, they do not completely close responsibility gaps found in the academic literature. Because of this, there are still concerns about how accountability would be divided in intricate collaborative operations including shared data, distributed autonomous systems, and multinational command structures.

The two case studies selected for this thesis are therefore not simply empirical illustrations but stress tests for the existing policy frameworks. The deployment of Merops counter-drones on NATO's eastern flank raises specific concerns about whether the Autonomy Implementation Plan and PRUs offer adequate guarantees of human oversight, explainability, and governability when autonomous engagement may take place in the context of intense electronic warfare and short decision timelines. The Lithuanian GA-10FPV-AI and Hornet drones, on the other hand, demonstrate how NATO and EU standards on effective human control, target discrimination, and civilian protection interact with innovation by a smaller ally and later transfer to Ukraine. In both situations, the policy analysis will look at whether the current combination of Alliance-level principles, EU soft norms, and national practices is sufficient to stop the deterioration of international security and military ethics, or if more extensive regulatory changes would be necessary to make the integration of AI into autonomous weapons truly compliant with the standards mentioned in the literature.

3.6 Case studies

The case studies chosen are two, in order to exemplify the incorporation of artificial intelligence (AI) into autonomous weapon systems. Merops Counter-Drone system and Lithuanian AI-Enabled Granta GA-10FPV-AI will be analyzed.

CASE 1: NATO DEPLOYMENT OF THE AI-ENABLED MEROPS COUNTER-DRONE SYSTEM

The Merops counter-drone system is an example of how AI is directly incorporated into operational autonomous weapon systems used by several NATO members. It also demonstrates a tangible NATO response to a rapidly changing threat environment, where low-cost autonomous drones challenge conventional air defense models. This example demonstrates how AI enhances autonomous targeting and decision-making in practical defense scenarios.

The U.S. created the Merops system (Project Eagle), a counter-unmanned aerial system (C-UAS) that combines autonomous interceptor drones (Surveyor) with radar, electro-optical, and infrared sensors. When satellite or radio links are jammed, the sensors identify and categorize approaching UAV threats, providing targeting information to human operators or enabling the interceptor drones to engage on their own. The system is useful against threats that take advantage of GNSS denial or other interruptions since it is built to operate under contested electronic warfare settings. Its interceptor drones demonstrate an innovative combination of autonomy and reusable strike capabilities. They can smash hostile drones or carry tiny warheads, and if they miss, they can use parachutes to recover (Complex Discovery, 2025).

NATO partners Poland, Romania, and Denmark started installing the Merops system along the alliance's eastern frontier in late 2025 to prevent and counter unmanned aircraft threats in response to repeated drone breaches into Polish and Romanian airspace. This deployment is a component of larger initiatives under NATO's Eastern Sentry/Eastern Flank Deterrence Line strategy to update its defenses against low-intensity airborne intrusions and hybrid threats. The cost-effectiveness of Merops in comparison to conventional air defense systems (such fighter jets and missile interceptors) and its AI-assisted autonomy to function even in the face of electronic warfare have been highlighted by NATO authorities (Wellman, 2025).

This thesis looks at two key aspects to analyze the Merops instance in a way that connects practice to the more general research question. It first examines how the technology alters NATO's defensive posture by providing more persistent counter-drone capabilities and what this means for Allied and partner forces on the eastern flank in terms

of readiness and cost-effectiveness. Second, it examines the moral and legal framework of Merops' deployment, posing the question of how its partially autonomous engagement decisions strike a compromise between NATO rules of engagement and international humanitarian law, especially with regard to human oversight, accountability, and the possibility of wrong targeting.

4.1 Technical and operational profile

To counter Russia's increasing deployment of unmanned aerial systems along the Alliance's eastern flank, some NATO partners have deployed Merops, a small, AI-enabled counter-drone device developed in the United States. The platform combines machine-learning algorithms with radar, electro-optical, and infrared sensors to identify, categorize, and track small drones that are challenging for traditional air-defense radars to detect. Once a target has been located, Merops can either transfer the track to other air-defense systems or fire small interceptor drones that physically neutralize or redirect the hostile UAV. This strategy is intended to be effective even in highly congested electromagnetic settings (Esgueva, 2026).

Merops is made to be extremely transportable and quickly deployable from an operational standpoint. As threat patterns change, the system may be swiftly deployed to safeguard airbases, border areas, or vital infrastructure because it is compact enough to fit in the back of a mid-size pickup truck. In keeping with lessons learned from Russian and Iranian-made drones that take advantage of holes in conventional air-defense coverage, its autonomous navigation and sensor-fusion capabilities are designed to continue operating when satellite navigation or communications links are compromised. The system is appealing to Allies facing comparable threats because early versions were battle-tested in Ukraine and reportedly neutralized a significant number of Shahed-type loitering munitions (Burrows, 2026).

According to publicly accessible descriptions, Merops functions in a "human-on-the-loop" arrangement when it comes to human-machine interaction. According to the taxonomy used in this thesis, Merops is therefore viewed as a primarily human-on-the-loop system: AI modules perform ongoing threat assessment and surveillance, make engagement recommendations, and can start intercepts within predetermined parameters,

while human operators maintain the formal authority to approve or reject engagements and to override the system when needed.

While human operators have formal ability to authorize or cancel interceptions, AI-based modules conduct ongoing surveillance and threat assessment, producing engagement recommendations. This architecture attempts to balance persistence and speed with some degree of human judgment, but it also implies that algorithms contribute a large portion of the situational framing, which affects the options accessible to commanders in real-world scenarios (Complex Discovery, 2025).

4.2 Impact on NATO’s defensive posture and readiness

NATO's defensive strategy on the eastern flank has started to change in at least three ways as a result of the deployment of Merops. First, the system allows for more continuous counter-UAS coverage than could be practically maintained by human operators alone by automating the identification and tracking of inexpensive drones, particularly in areas like Poland and Romania that frequently experience Russian drone incursions close to their borders. This is in line with Allied intentions to use loitering munitions and other low-cost aircraft threats to prevent and repel cross-border harassment by maintaining "always-on" vigilance over crucial links (Burrows, 2025).

Second, Merops helps make NATO's air defense more affordable and scalable. The system provides a more economically viable response to small, slow, and numerous enemy UAVs because it employs relatively cheap interceptor drones instead of pricey surface-to-air missiles, each of which can be significantly less expensive than conventional interceptors. For Allies with tight defense budgets, this cost asymmetry is crucial because they must protect long frontiers without depleting pricey missile supplies. The capacity to gradually improve Merops' software and classification models is also consistent with Alliance initiatives to "spiral" capability development instead of depending exclusively on significant hardware initiatives (Burrows, 2026).

Third, NATO's transition to layered, AI-enabled air defense infrastructures is strengthened by the deployment of Merops. The system is envisioned as one layer inside larger concepts like Operation Eastern Sentry and other eastern-flank reinforcement measures, and it is being incorporated into national and Alliance command-and-control organizations in Poland, Romania, and Denmark. This pushes NATO closer to a defense

posture in which human commanders rely more and more on computational technologies to prioritize threats, sift sensor data, and suggest large-scale actions. This increases NATO's reliance on intricate software systems whose behavior can be challenging to forecast or audit, particularly in the context of electronic warfare, even while it may improve preparedness and resilience against saturation attacks (Malyasov, 2025).

4.3 Ethical and legal context of autonomous engagement

The Merops case immediately calls into question how high-tempo, AI-enabled air defense can uphold NATO's commitment to meaningful human control and adherence to international humanitarian law (IHL). The Alliance rules of engagement, which outline when and how lethal force may be used in peacetime air policing and collective defense scenarios, as well as the same jus in bello requirements of distinction and proportionality that apply to other uses of force, essentially govern counter-drone operations. Accurate categorization and trustworthy human oversight are essential to preventing incorrect engagements because Merops is mostly used in border regions where commercial drones, allied military platforms, and civilian air traffic may share airspace.

The six Principles of Responsible Use (lawfulness, responsibility, explainability, reliability, governability, and bias mitigation) are outlined in NATO's revised 2024 Artificial Intelligence Strategy at the doctrinal level as the foundation for all AI-enabled capabilities. The 2022 Autonomy Implementation Plan and the Data and AI Review Board (DARB), which work together to guarantee that AI and autonomous systems continue to be subject to human judgment, legal scrutiny, and post-hoc accountability, complement these principles. Therefore, in theory, Merops should be built and run in a way that allows a human operator to comprehend the system's suggestions, interfere with its operation, and be held accountable for engagement choices.

In reality, the margin for thoughtful human judgment may be compressed by the speed and complexity of counter-drone actions. Operators may only have a few seconds to authorize the launch of an interceptor when algorithms carry out continuous surveillance, classification, and threat assessment. They may also rely significantly on the system's confidence indicators rather than independently confirming each track. Particularly under situations of exhaustion or saturation, this dynamic runs the risk of shifting from truly meaningful human control to a merely nominal "human in the loop"

who just rubber-stamps machine outputs. From an ethical standpoint, this supports the literature's worries that AI-driven targeting can weaken operators' sense of moral obligation and obfuscate culpability in the event that a civilian or affiliated platform is improperly involved.

Merops's multi-stakeholder structure makes accountability even more challenging. A U.S. defense contractor created the system, which is then implemented by specific Allies like Romania and Poland and incorporated into NATO's larger air-defense framework. It is unclear whether the operator, the national chain of command, the developer, or NATO's collective institutions would bear primary responsibility if a misidentification results in unlawful harm, such as the destruction of a civilian drone that causes casualties on the ground. Joint operations utilizing Merops may result in responsibility gaps noted in the academic literature on autonomous weapons because current Alliance documents mainly defer legal obligation to member nations.

Lastly, the system may have an impact on escalation management and crisis stability. Merops contributes to a forward, highly automated defense plan against Russian drones and, indirectly, against their launch platforms since its goal is to make NATO's eastern flank "so well-armed that Moscow's forces will be deterred from ever contemplating crossing." This could increase deterrence, but it also increases the risk of quick, automatic encounters close to delicate borders, where misclassification of an unclear UAV or human aircraft could lead to diplomatic problems or military escalation. These hazards are only mentioned in broad strokes in current NATO policy documents, which place more emphasis on interoperability and technology advantage than on specific operational measures to prevent unintentional escalation in human-machine teams.

4.4 Assessment against NATO and EU governance criteria

The Merops example seems normatively promising but practically unclear when compared to the four criteria established in this thesis: meaningful human control, accountability, compliance with IHL and human rights, and consequences for deterrence stability. The system is subject to national weapons-review procedures and the Alliance's Principles of Responsible Use, and it can theoretically be incorporated into chains of command that maintain human decision-making authority. However, NATO's current governance framework does not fully address the ethical and security concerns raised by

AI-enabled air defense, as evidenced by the heavy reliance on AI-driven classification under time pressure, the fragmented allocation of accountability between states and contractors, and the scant attention paid to escalation dynamics.

CASE 2: LITHUANIAN AI-ENABLED GRANTA GA-10FPV-AI

A smaller NATO member's domestic innovation in AI drone technology is demonstrated by the Lithuanian Granta GA-10FPV-AI and related autonomous drones (such as Hornet XR), which have significant implications for both national defence and coalition operations. As part of a larger trend in distributed autonomous operations, these UAVs show how AI autonomy is being employed not only for detection and interception but also for navigation and mission execution under difficult conditions.

Granta Autonomy in Lithuania created the GA-10FPV-AI, a first-person-view (FPV) quadcopter with autonomous navigation capabilities for reconnaissance and loitering tasks. These drones' capacity to do pre-programmed tasks and function even in the absence of GNSS or GPS signals is made possible by AI, which is crucial in contested areas where adversaries may utilize spoofing or jamming. The VTOL (vertical takeoff and landing) capabilities of the GA-10FPV-AI improves operational flexibility in a variety of urban and terrain conditions. The Lithuanian Armed Forces are purchasing thousands of these drones and donating them to Ukraine, where they have been used on the battlefield and continuously improved under combat conditions (Gosselin-Malo, 2024).

This is another NATO-aligned autonomous UAV used for intelligence, surveillance, and reconnaissance (ISR) tasks, which includes Hornet XR. These drones are useful for front-line intelligence and targeting data because they are small, hand-launched, and able to function even in GPS-denied or electronically challenged areas.

This thesis focuses on two primary dimensions to analyze the Lithuanian GA-10FPV-AI drones as a second case study. By analyzing how NATO allies gain from shared autonomous systems and from the integration of data and operational experience across forces, it first looks at the implications for cooperative operations, including aid to Ukraine. Second, it examines the normative ramifications of using autonomous strike and reconnaissance UAVs for target discrimination, civilian risk, and adherence to international law, as well as the strategic and ethical issues brought up by Lithuania's participation in the Alliance's developing autonomous weapons ecosystem.

5.1 Technical and operational profile

The Lithuanian company Granta Autonomy created the GA-10FPV-AI, a first-person view (FPV) quadcopter drone intended for military uses like close-range reconnaissance and precision strikes. The system can strike armored vehicles, artillery pieces, and fortified positions at the tactical level because it can carry a warhead weighing up to 3 kg and has an operational range of about 20 kilometers. Its distinguishing feature is an AI-enabled guiding and anti-jamming suite. The drone's onboard algorithms are made to maintain control and lock-on capabilities even in the face of intense electronic warfare, rather than just switching between frequencies to avoid interference (Martyniuk, 2025).

As part of a larger national plan to build indigenous drone capabilities and bolster Kyiv's defense, Lithuania contracted with Granta Autonomy in 2024–2025 to supply thousands of GA–10FPV–AI drones to both Ukraine and its own military forces. The drones have been tested and used in the high-intensity military environment of Ukraine, where frequently moving front lines and thick Russian jamming offer a challenging testing ground for autonomous and semi-autonomous systems. The GA-10FPV-AI's AI components are said to help with target tracking, navigation, and terminal guiding, blurring the distinction between remote-controlled and autonomous attack capabilities, even though operators still initiate missions and usually designate targets via FPV control (Global Defence News, 2025).

The GA-10FPV-AI is best described as a human-in-the-loop system with expanding human-on-the-loop features fm the standpoint of the human-in/on/out-of-the-loop taxonomy: human operators determine when and against which targets the drones are used, but AI-enabled guidance and anti-jamming functions increasingly influence how the mission develops after it is launched.

The GA-10FPV-AI provides a helpful perspective on how European Allies are experimenting with autonomy just outside NATO's official theater of operations, given Lithuania's dual role as a NATO front-line state and a supplier of AI-enabled drones to a partner at war. It is conceivable that future Alliance doctrine and procurement, including the possible application of comparable equipment in collective defense situations, will be influenced by operational lessons learned from Ukraine.

5.2 Impact on deterrence, doctrine and alliance posture

NATO-relevant security dynamics are impacted by the Lithuanian GA-10FPV-AI program in a number of ways. First, Lithuania and other European donors hope to counterbalance Russian might and support a deterrence-by-denial stance on NATO's eastern frontier by providing Ukrainian forces with a significant number of reasonably priced, AI-assisted attack drones. Such systems are appealing for saturating enemy locations while limiting threats to friendly personnel because to their precision, anti-jamming robustness, and expendability. This is consistent with larger trends in Alliance thinking toward unmanned, attritable platforms (Hambling, 2025).

Second, a doctrinal change toward integrating AI into its own tactical forces is indicated by Lithuania's domestic use of GA-10FPV-AI drones. Lithuania has a strong motivation to build nimble, networked strike capabilities that can thwart any possible aggressor's plans because it is a small ally situated close to Belarus and the Russian outpost of Kaliningrad. If AI-enabled FPV drones are successful in Ukraine, there will probably be more pressure on NATO to incorporate comparable systems into ideas and exercises for collective defense, including scenarios in which Allied troops might use them on Alliance territory in accordance with NATO rules of engagement. As a result, there is a feedback loop whereby wartime experimentation outside of NATO's official mandate shapes the Alliance's stance going forward (Allison, 2025).

Third, worries regarding proliferation and arms race dynamics are highlighted by the broad availability of GA-10FPV-AI systems. The threshold for launching targeted assaults may be decreased when more states and non-state actors obtain inexpensive, AI-assisted loitering bombs. This could have repercussions for crisis stability and escalation management in the larger Euro-Atlantic region. This supports the larger body of literature's caution that AI-driven autonomy can hasten hostilities and make it more difficult for political leaders to maintain real control over the use of force once hostilities have started.

5.3 Ethical and legal context

The GA-10FPV-AI poses comparable ethical and legal issues as Merops, but in a more offensive, strike-oriented setting. Because the drone is intended to deliver deadly payloads in intricate, frequently urban settings, both the human operator's discretion and

the accuracy of the AI-assisted targeting features are necessary to ensure compliance with distinction and proportionality. There is a chance that misclassification or training-data bias could result in illegal strikes on civilian targets that resemble valid military targets if machine-learning models direct terminal approach or auto-lock onto visually recognized targets.

IHL and, in many situations, national laws mandating "meaningful human control" over the use of lethal force continue to bind Lithuania and other European nations who purchase or export such systems. However, other from broad remarks about human accountability and legal scrutiny, neither NATO nor the EU now offer a comprehensive, legally enforceable definition of what meaningful human control for small, swarming drones implies. Similar to Merops, this allows for different national interpretations of what level of explainability is necessary for post-strike assessments, how closely operators must oversee AI-assisted activities, and how much autonomy is allowed in terminal guidance.

The EU's Artificial Intelligence Act, which prohibits military applications while still influencing dual-use industry standards, also creates a governance gap, as demonstrated by the GA-10FPV-AI case. For their civilian AI solutions, Granta Autonomy and comparable companies must adhere to EU standards for transparency, risk management, and human oversight; nevertheless, their military offers are subject to national defense restrictions, which may be less comprehensive or standardized. It is more difficult to guarantee uniform ethical standards among European suppliers of AI-enabled weaponry to NATO members and allies because of this disjointed system.

5.4 Assessment against NATO and EU governance criteria

The Lithuanian GA-10FPV-AI drones demonstrate the possibilities and dangers of AI-assisted lethality in a NATO-relevant setting when applied to the four criteria of this thesis. On the one hand, Lithuanian shipments to Ukraine take place inside a political context that openly emphasizes adherence to IHL and responsible use, while operators continue to plan and oversee assaults, maintaining a formal layer of human control. However, it may be challenging to ensure meaningful human control and transparent accountability in practice due to the growing autonomy of guidance and anti-jamming functions, the severe operational demands of the Ukrainian battlefield, and the lack of

Alliance-wide or EU-wide binding standards for small autonomous strike systems. Similar to Merops, these ambiguities imply that NATO's existing governance system is operationally under-specified but normatively ambitious, especially when Allies create and use AI-enabled weapons outside of the Alliance's official command structures.

DISCUSSION

In the following section, the MEROPS counter-drones system and the Lithuanian GA-10FPV-AI drones are compared against the four criteria which have been developed in the thesis (meaningful human control, accountability, compliance with international humanitarian and human rights law, and implications for deterrence stability and escalation).

6.1 Typology of risks and safeguards

The two case studies highlight a similar pattern of dangers and protections related to AI-enabled weapons in NATO and its surrounding areas. In terms of risk, both Merops and the GA-10FPV-AI demonstrate how growing autonomy can undermine significant human control under time constraints, distribute accountability among several actors, and open up new avenues for escalation in congested, electronically disputed contexts. Simultaneously, they show how AI can improve force protection, accuracy, and resilience, particularly when systems are integrated into formal rules of engagement, subject to weapons reviews, and run with clear human-in/on-the-loop criteria. This implies that ethical and legal norms are not intrinsically incompatible with AI integration; rather, the balance of incentives inside NATO currently favors performance and reactivity above restraint and transparency in the absence of strong, enforced protections.

6.2 NATO–EU governance tensions

The examples also draw attention to conflicts between the EU's and NATO's policies regarding autonomous weapons. The Principles of Responsible Use, the Autonomy Implementation Plan, and the Data and AI Review Board serve as the foundation of NATO's non-binding, principles-based framework, which establishes strict normative standards but mostly leaves implementation up to member states. The EU's Artificial Intelligence Act, on the other hand, establishes specific requirements for high-risk civilian AI systems while specifically excluding defense and national security applications, leaving military AI to rely on a variety of state regulations and political commitments.

The GA-10FPV-AI, manufactured in an EU and NATO member state but exported to a non-NATO war effort, and Merops, fielded by NATO Allies, fall precisely in this

governance gap: they are shaped directly by NATO principles and indirectly by EU standards, but they are not governed by a single, legally binding European regime. It is more difficult to guarantee uniform baselines on matters like testing, openness, human supervision requirements, and incident reporting throughout the Alliance as a result of this fragmentation.

6.3 Answers to the research sub-questions

The three sub-questions that guide this thesis can have provisional solutions thanks to the combined policy and case-study analysis. First, while AI-driven autonomy in weapons undoubtedly challenges conventional notions of accountability and meaningful human control, it does not completely eradicate them. Although humans are still formally responsible in both situations, algorithmic framing and operational tempo limit their capacity to make timely, well-informed decisions.

Second, while strengthening NATO's defensive posture and reducing the cost of precision strikes, AI-enabled systems have unclear effects on deterrence stability and escalation risk. They also shorten decision times and raise the possibility of misperception or unintentional escalation in high-tension situations. Third, the regulatory frameworks of the EU and NATO address ethical issues in a way that is both normatively ambitious and operationally inadequate. While they express strong principles of human control and responsible use, they do not have common, enforceable standards for specific systems such as Merops and GA-10FPV-AI, which leaves a great deal of room for different national interpretations and the emergence of responsibility gaps.

6.4 Meaningful human control

In practice, Merops is more akin to a "human-on-the-loop" paradigm, whilst the GA-10FPV-AI is still primarily operator-driven, but both systems formally maintain a human role in deadly decisions. Continuous surveillance, classification, and threat scoring are carried out by AI modules in Merops, giving operators short reaction times and a strong incentive to rely on system recommendations, which runs the risk of veering toward rubber-stamping rather than substantial judgment. While AI-assisted navigation and anti-jamming features progressively influence how targets are approached and engaged, GA-10FPV-AI operators still initiate and oversee individual strikes. All things considered, both examples indicate that NATO's focus on meaningful human control is

only partially operationalized and that the degree of human involvement can differ greatly between offensive and defensive AI-enabled systems (Burrows, 2026).

6.5 Accountability and responsibility

The distribution of blame for illegal harm is nonetheless disjointed in both situations. Although NATO command institutions, host-nation military, and U.S. developers are all involved in Merops, existing Alliance documents mainly shift legal responsibility to individual states, leaving it unclear who would be held ultimately responsible for a misidentification incident in a joint operation. The GA-10FPV-AI drones draw attention to a separate accountability issue: as Lithuania both runs the system and exports it to Ukraine, both the producing ally and a partner engaged in a high-intensity conflict bear some of the blame for any abuse or mistake. Scholarly worries over accountability gaps in autonomous weapons governance are reinforced by the fact that neither the current NATO nor EU frameworks offer comprehensive, enforceable guidelines for tracking decision-making via intricate human-machine interactions.

6.6 Compliance with IHL and human rights

Merops and GA-10FPV-AI are supposed to be employed in compliance with national and Alliance rules of engagement as well as IHL principles of distinction and proportionality. However, both systems function in areas where accurate target detection is challenging, such as dense, frequently urban battlefields for GA-10FPV-AI and border regions with mixed civilian and military airspace for Merops. Although AI-based classification and guiding can lessen the effort for humans, it also introduces opacity and potential bias, making it more difficult to reconstruct the reasons behind an object's identification as a legal target. It appears that adherence to IHL and human rights norms still largely rests on national practice rather than a strong common framework because there are no Alliance-wide, legally enforceable criteria of permissible autonomy or explicit transparency requirements for small, AI-enabled systems (Esgueva, 2026).

6.7 Deterrence stability and escalation risk

Lastly, the two examples show how autonomy afforded by AI can both increase deterrence and generate new escalation pressures. Merops contributes to a deterrence-by-denial strategy by strengthening continuous air defense and discouraging Russian drone

incursions near NATO's eastern border. However, highly automated encounters close to sensitive borders raise the possibility that technical issues or misinterpretation could lead to unintentional military escalation or diplomatic crises. In addition to supporting Ukraine's defense and demonstrating Lithuania's commitment to NATO, the GA-10FPV-AI drones contribute to the wider spread of low-cost, AI-assisted strike systems that could reduce the threshold for deploying force and quicken the pace of combat. The incidents collectively reinforce the literature's caution that the pursuit of resilience and efficiency in AI-enabled systems may surpass attempts to control escalation and maintain strategic stability in the absence of clear, operational safeguards (Gosselin-Malo, 2024).

When considered collectively, the comparative analysis shows that the incorporation of AI into autonomous and semi-autonomous weapons in the NATO context is conditionally compatible with ethical and security principles: formal commitments to human control and legality exist, but their implementation is uneven, and existing governance frameworks leave a great deal of room for emerging responsibility gaps and divergent national interpretations.

CONCLUSION

The purpose of this thesis was to assess to what extent, in the context of NATO, the integration of artificial intelligence into autonomous weapons challenges military ethics and international security. It looked at how important ideas like meaningful human control, accountability, adherence to international humanitarian and human rights law, and deterrence stability are understood and used in practice using a qualitative methodology based on literature review, policy analysis, and two illustrative case studies. The analysis concentrated on the deployment of Lithuanian GA-10FPV-AI drones and the Merops counter-drone system in and around the Alliance, as well as the EU's developing AI regulatory framework and NATO's changing governance of AI and autonomy.

The results imply that while AI-enabled autonomy presents significant difficulties, it does not inherently exclude morally and legally acceptable warfare. Although humans are formally involved in the decision-making process in both case studies, algorithmic framing and operational tempo limit their capacity to make timely, well-informed decisions, particularly in high-pressure air defense and combat situations. Additionally, accountability is still dispersed: developers, national chains of command, and NATO structures share accountability for possible illegal harm, and current frameworks provide little means for tracking and assessing intricate human-machine interactions after the fact.

At the level of principles, compliance with IHL and human rights norms is strongly affirmed; however, specific guidelines regarding acceptable levels of autonomy, transparency requirements, and incident reporting are still lacking, especially for small, dispersed systems like FPV strike drones and counter-drone platforms. The study concludes that, in terms of international security, AI-enabled technologies simultaneously improve NATO's risk management and defensive posture while speeding up decision-making and igniting arms race dynamics that, if unchecked, might jeopardize crisis stability.

In general, the thesis contends that the incorporation of AI into autonomous and semi-autonomous weapons within the NATO framework is only marginally consistent with international security and military ethics. Strong, verifiable methods for genuine human control, distinct lines of accountability, strong legal review, and specific

precautions against unintentional escalation are all necessary for compatibility, but present governance only partially and fragmentarily offers these elements. Together with the EU's wider normative influence, NATO's Principles of Responsible Use and Autonomy Implementation Plan articulate lofty ideals but do not establish a single, legally enforceable European defense AI regime. Systems like Merops and the GA-10FPV-AI are being developed and implemented in this gap.

Numerous policy recommendations follow from these findings. First, NATO should go beyond high-level ideals and establish a more operational definition of meaningful human control, distinguished by system type and degree of autonomy, so that Allies have clear expectations on human roles in detection, targeting, and engagement. This directly addresses the Merops case, where human-on-the-loop arrangements under extreme time pressure run the risk of becoming merely nominal oversight. Second, the Alliance and its members should establish uniform procedures for documenting and looking into AI-related incidents. These procedures should include minimum standards for recording algorithmic decisions and providing pertinent data for post-hoc legal and ethical review. This is because in both cases, it is challenging to reconstruct the specifics of individual engagements due to the fragmented responsibility among developers, Allies, and partners.

Third, the Lithuanian GA-10FPV-AI program, which operates at the intersection of EU export controls, national defense policy, and NATO's principles of responsible use, serves as an example of how NATO and the EU should collaborate more closely on minimum standards for defense-AI testing, transparency, and human-oversight requirements in order to prevent a regulatory gap between civilian and military uses of similar technologies, even though the AI Act expressly excludes military applications. Lastly, since it is most difficult to ensure accurate distinction and proportionality in dense, dynamic environments, Allies should consider agreeing on red lines or restrictions for fully autonomous lethal systems in populated areas, reflecting the broader concern that highly automated engagements in complex theatres, such as Ukraine or NATO’s eastern flank, could trigger disproportionate harm or inadvertent escalation.

It is important to recognize the limits of this thesis. The analysis does not include classified material or internal NATO paperwork; instead, it is solely based on publicly accessible sources, such as academic research, policy documents, and media reporting.

Additionally, there may be linguistic bias because the majority of the primary and secondary materials studied are in English and, to a lesser extent, Spanish. As a result, opinions and technical details published in Lithuanian, Polish, Romanian, or Russian may be underrepresented. Furthermore, interviews with system developers, military operators, or officials directly involved in the deployment of Merops or GA-10FPV-AI have not been possible, which limits the capacity to confirm how governance frameworks are implemented in daily operations. Lastly, some operational numbers claimed in open sources, like the number of Shahed drones purportedly neutralized by Merops or the precise amounts of GA-10FPV-AI supplied, should be regarded as indicative rather than final since they cannot be independently validated.

These limitations also point to areas that need more research. Access to confidential operational data, structured interviews with practitioners, and a larger range of AI-enabled systems beyond the two case studies analyzed here would be beneficial for a more thorough evaluation of how frequently responsibility gaps and control issues occur in practice. Future studies could also examine how various alliance structures (such as ad hoc coalitions or EU-only missions) interpret and apply principles pertaining to autonomy, as well as how new arms control initiatives may limit or alter the development of AI-powered weapons in Europe.

BIBLIOGRAPHY

Allison, G. (2025). *Lithuania's Granta unveils X-WING loitering munition*.

Amanda Sharkey. (2019). Autonomous weapons systems, killer robots and human dignity.

ARROWS. (2025). Legal aspects of the development of weapon systems with artificial intelligence in 2025. https://arws.cz/news-at-arrows/legal-aspects-of-the-development-of-weapon-systems-with-artificial-intelligence-in-2025

Badreldin Hamad. (2025). The Impact of Digital Technology on International Humanitarian Law: Ethical and Legal Implications of Autonomous Weapons Systems. *European Journal of Law and Political Science,* https://www.ej-politics.org/index.php/politics/article/view/182/220

Brian Stauffer. (2025). *A Hazard to Human Rights Autonomous Weapons Systems and Digital Decision-Making.* https://www.hrw.org/report/2025/04/28/a-hazard-to-human-rights/autonomous-weapons-systems-and-digital-decision-making

Burrows, E. (2025). *A new system to identify and take down Russian drones is being deployed to NATO's eastern flank*. https://www.ap.org/news-highlights/spotlights/2025/a-new-system-to-identify-and-take-down-russian-drones-is-being-deployed-to-natos-eastern-flank/

Burrows, E. (2026). *The U.S. is sending an AI-powered anti-drone system to the Mideast as response to countering Iran's Shahed has been 'disappointing.'* https://fortune.com/2026/03/07/us-anti-drone-system-merops-mideast-iran-shahed/

Complex Discovery. (2025). NATO's Counter-Drone Solution: Inside the Merops Deployment Reshaping European Air Defense.

Daniele Amoroso, &., & Guglielmo Tamburrini. (2020). *Autonomous Weapons Systems and Meaningful Human Control: Ethical and Legal Issues*

David Arias Guedón, , & Javier García Luengo. (2024). La inteligencia artificial militar ha sido excluida del Reglamento europeo de IA, pero la OTAN y la normativa española han abordado estos usos. *Garrigues,* https://www.garrigues.com/es_ES/garrigues-digital/inteligencia-artificial-militar-ha-sido-excluida-reglamento-europeo-ia-pero-otan

Dr. Nadia Burkart, D. M. A.,& & Fabian Jeschke. (2025). Ethical Considerations for the Military Use of Artificial Intelligence in Visual Reconnaissance.

Elisabeth Gosselin-Malo. (2024). Lithuania to deliver thousands of FPV drones to its army, Ukraine.

Esgueva, M. (2026). *U.S. MEROPS: Counter Drone System for NATO's Eastern Flank*. https://greydynamics.com/us-merops-counter-drone-system-nato-eastern-flank-romania-poland/

Ethan Lee. (2025). What are Autonomous Weapon Systems? *Harvard,* https://www.belfercenter.org/what-are-autonomous-weapon-systems

European Union. (2021). Guidelines for military and non-military use of Artificial Intelligence. https://www.europarl.europa.eu/news/en/press-room/20210114IPR95627/guidelines-for-military-and-non-military-use-of-artificial-intelligence

Hambling, D. (2025). *Battlefield Experience And Advanced Software Build Better Drones*. *Forbes*. https://www.forbes.com/sites/davidhambling/2025/03/18/battlefield-experience-and-advanced-software-build-better-drones/

Jie Guo. (2025). The ethical legitimacy of autonomous Weapons systems: reconfiguring war accountability in the age of artificial Intelligence. https://www.tandfonline.com/doi/full/10.1080/16544951.2025.2540131

Johnson Nancy, F. W., Rose Hoover. (2025). AI-Driven Drone Swarms and Their Implications for Future Battlefields.

Malyasov, D. (2025). *NATO fields Merops counter-drone system on eastern flank*. https://defence-blog.com/nato-fields-merops-counter-drone-system-on-eastern-flank/

Mariarosaria Taddeo & Alexander Blanchard. (2022). A Comparative Analysis of the Definitions of Autonomous Weapons Systems.

Martyniuk, Y. (2025). *Lithuania sends Ukraine cutting-edge AI drones that don't just avoid Russian jamming—they overcome it*. https://euromaidanpress.com/2025/03/18/lithuania-sends-ukraine-cutting-edge-ai-drones-that-dont-just-avoid-russian-jamming-they-overcome-it/

Mengkorn Pum. (2024). Autonomous Weapons Systems: Ethical Concerns and International Regulation in the Use of AI in Military Applications.

Mercy Abiola Gilbert & Chris Gilbert, (2024). The Security Implications of Artificial Intelligence (AI)-Powered Autonomous Weapons: Policy Recommendations for International Regulation.

NATO. (2017). Autonomous Military Drones: No Longer Science Fiction.

News, G. D. (2025). *Lithuania to Supply Thousands of GA-10-FPV-AI Drones to Ukraine*. https://www.armyrecognition.com/archives/archives-land-defense/land-defense-2024/lithuania-to-supply-thousands-of-ga-10-fpv-ai-drones-to-ukraine

Summary of NATO's Autonomy Implementation Plan, (2022). https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2022/10/13/summary-of-natos-autonomy-implementation-plan

Ozlem Ulgen. (2017). *Kantian Ethics in the Age of Artificial Intelligence and Robotics*. https://www.qil-qdi.org/wp-content/uploads/2017/10/04_AWS_Ulgen_FIN.pdf

Seth Lazar. (2017). Just War Theory: Revisionists vs. Traditionalists. https://philpapers.org/archive/LAZJWT.pdf

Summary of NATO's revised Artificial Intelligence (AI) strategy, (2024). https://www.nato.int/en/about-us/official-texts-and-resources/official-texts/2024/07/10/summary-of-natos-revised-artificial-intelligence-ai-strategy

Phillip Walter Wellman. (2025). After drone breaches, NATO turns to Merops system for eastern flank defense.

Stuart Casey-Maslen. (2025). Autonomous Weapons Systems Under International Law. https://doi.org/10.1007/978-3-031-95578-5_10

Universidad Francisco de Vitoria. (2025). Is international security and global security the same? *Universidad Francisco De Vitoria,* https://www.ufv.es/is-international-security-and-global-security-the-same-preguntas-masters/

ANNEX: Declaration of Use of Generative AI Tools

Academic Year: 2025-2026
Master's Programme: Master in International Security Management (MISM)
Student Name: Clara Núñez Vázquez

I declare that generative artificial intelligence tools have been used as support tools in the preparation of this Master's Final Thesis.

☐ x YES ☐ NO

1. Ethical and Academic Use

Have you included sensitive or personal data when using AI tools? If yes, specify:

_____
_____

Have you used AI tools to replace your own work without critically reviewing the generated content? If yes, specify:

_____
_____

Have you followed the academic recommendations and guidelines regarding the use of AI tools?

__Yes_____
_____

2. Technical Use of AI Tools

Please indicate the AI tools used (e.g., ChatGPT, Copilot, Claude, Gemini):

Claude

Please mark the applicable uses:

- ☐ Text generation
- ☐ Reformulation / editing
- ☐ Translation / proofreading
- ☐ X Structure suggestions
- ☐ Methodological support
- ☐ Bibliographic search or citation support

- ☐ Audiovisual content generation
- ☐ Other uses (please specify)

I confirm that the final content of this thesis has been fully reviewed, corrected, and validated by me as the author. The use of AI has not replaced my own critical analysis, personal reflection, or intellectual work.

Signature: Clara Núñez Vázquez