Artificial Intelligence, Drones and Cyber Warfare: Ethical and Anticipated Ethical Issues

Richard L Wilson[1] and Noah Donnelly[2]
[1]Department of Philosophy/Computer Science and Information Sciences, Towson University, Baltimore, Maryland, USA
[2]Computer Science and Information Sciences, Towson University, Baltimore, Maryland, USA

wilson@towson.edu
ndonnel1@students.towson.edu

Abstract: Artificial Intelligence plays an active role when both drones and Cyber Warfare are introduced into kinetic wars. The importance of drones for kinetic warfare points to the importance of cyber warfare for countering drone warfare. A drone that flies, reroutes, and avoids threats without human interaction no longer operates as a remote-controlled vehicle. Instead, it acts autonomously in combat situations. Small quadcopters in Ukraine were equipped with AI navigation packages, enabling them to complete their missions even after Russian troops had jammed their GPS signals. Programs in the US, such as Darpa Evade, are deploying Sikorsky's autonomy MATRIX software in vertical takeoff drones, which nominally have the capabilities of flight and fight without communication with an operator, while they have situational awareness, and continue to achieve their objectives. The most contentious application of AI in targeting military drones is the recognition of targets. Using deep learning models trained on massive amounts of image data, drones can recognize vehicles, artillery, or even people. The AI models utilize visual patterns and infrared signatures to help identify and distinguish between an enemy armored vehicle on a battlefield, hidden in a field, or a civilian tractor, also hidden in the same spot. Russia's KUB loitering munition features an AI visual identification module designed to enhance its search range and accuracy. The Bayraktar TB2 features onboard targeting systems that provide unified target designation and tracking. The TB2's follow-on munition, the Kamankes loitering munition drone, is advertised to have AI-guided optics that enable it to loiter, identify, and autonomously kill any targets (Aviacionline, 2023; Baykar Technology, 2026). A drone that could perform all of these activities could loiter for hours, detect a missile launcher or armor target, and kill without verification by human input at any point. Ukraine's advances in AI are even more unprecedented when factoring in that they are using modular AI kits to sever the "last mile" of the target kill chain, reducing the time from hours to minutes. AI proponents argue that accuracy enables speed and precision with minimal risk of unintended consequences. Critics caution that algorithmic-driven lethal decisions can go wrong if an object is incorrectly identified, leaving the laws of armed conflict with challenging and fuzzy questions of moral, legal, and professional liability. This analysis is concerned with identifying how AI can be applied in the context of cyberwarfare to counter the capabilities of AI being employed to enhance the capabilities of Drones as they are currently employed in military operations. In addition, this analysis will focus on ethical and anticipated ethical issues related to how AI can be applied in the context of cyberwarfare operations to counter the capabilities of AI being employed to enhance the capabilities of Drones as they are currently employed in military operations.

Keywords: Artificial intelligence, Unmanned aerial vehicles, Cyber warfare, Autonomous weapons systems, Anticipatory technology ethics, Adversarial machine learning, Meaningful human control

1. Introduction

Drones have historically operated primarily as remote-controlled vehicles, relying on continuous telemetry, data links, and human in the loop decision-making. Within today's battlespace, we have witnessed a forceful and irrevocable transition from automated systems to fully autonomous agents. An unmanned aerial vehicle (UAV) that flies, re-routes, avoids threats, and detects targets without any human interaction is no longer an extension of a human pilot, it is now functioning as an independent combat entity in a constantly changing battle environment (Wilson & Donnelly, 2026a). This change influences the strategic and tactical environment for global war.

Conventional telecom structures are susceptible to degradation under extremely competitive conditions including massive electronic warfare (EW), radio frequency interference, and signal jamming technologies. Military agencies are increasingly adding onboard AI-based navigation packages and advanced computer vision modules to small quadcopters and fixed-wing UAVs to help sustain their operational capabilities. These edge computing systems rapidly shorten the "last mile" from the target kill chain; as a result, the time between target identification and lethal engagement is significantly reduced from hours to minutes or seconds (Cybernews, 2026). The interactive relationship between the convergence of these technologies is transforming cyber warfare as well. Autonomous UAVs function as highly integrated nodes inside large, interrelated cyber physical systems and the Internet of Military Things (IoMT). The complex artificial neural networks and algorithms used to conduct the activities of autonomous drones provide adversaries with previously unseen cyber vulnerabilities, such as the potential use of adversarial machine learning, sensor spoofing, data poisoning, and protocol-level deception. Consequently, safe cyber warfare operations are increasingly focused on taking advantage of the cognitive and perceptual layers of autonomous drones by hijacking, misdirecting, or disabling these systems prior to arriving at their intended target (Erpek et al., 2024; Xi et al., 2025).

There is a high degree of uncertainty due to how cyber operations interact with kinetic operations, which raises issues related to the definition of war in terms of international law and escalation matrices for conflicts (Wilson & Donnelly, 2026a). The proponents of these technologies point to how AI increases speed of execution, accuracy and robustness while minimizing the chance for inadvertent actors, however; using algorithmic methods to determine whether an individual attack will be lethal is a significant ethical, legal and philosophical dilemma. In the event of incorrectly identifying a target (due to biases within algorithms, contextual variables within the environment or deceptive actions from an adversary), unintentional consequences can be the result. This poses significant challenges to the fundamental principles of the Laws of Armed Conflict (LOAC) (Wilson & Donnelly, 2026a). To address these challenges, this analysis presents empirically-based discussion of the technical mechanisms of AI deployment within military drones, an exploration of the technical attack surface of AI-driven (cyber) vulnerabilities within military drones, a comparative analysis of four real-world case studies involving AI-driven decision-making within military drones, as well as an application of the ACM Code of Ethics and the Anticipatory Technology Ethics (ATE) framework with respect to AI-driven military drones. The analysis concludes with specific technical and governance recommendations directed toward establishing a system of laws that will facilitate safe navigation through a rapidly evolving environment regarding autonomous algorithmic warfare.

2. Technical Issues

With military drones, Artificial Intelligence (AI) will be used to create highly sophisticated systems incorporating multiple integrations beyond just hardware and software across many applications like autonomous navigation; advanced computer vision; and cyber-resilient communication systems/architecture. These technologies are constantly evolving with new capabilities being introduced every day. As they become functional as weapons systems, they present new attackable surfaces that will increase the number of sophisticated cyber vulnerabilities, making it difficult to differentiate between a digital attack and a physical attack.

2.1 Autonomous Navigation in GPS-Denied Environments

GPS is an important aspect of traditional UAV operations and provides accurate real-time position, velocity, and altitude data of UAVs both before and during flight. GPS and other sensor information are incorporated into the UAV's flight controller via complex proportional-integral-derivative (PID) control loops. The position control loop uses the current coordinate and the desired coordinate of the UAV to identify the velocity at which the UAV should fly to achieve its desired position in three-dimensional space. The velocity control loop employs the velocity output from the position control loop to determine how much thrust the motors of the UAV must produce to achieve that velocity. The velocity output also compensates for external factors affecting the UAV's ability to achieve that velocity, such as wind drag (Florida International University, 2024).

However, many of today's adversaries use advanced GPS spoofing and broadband electronic jamming technologies to disrupt the link between the UAV and the satellite constellation that provides GPS data to drones. To protect against these vulnerabilities, increasingly sophisticated UAVs are using AI-enabled autonomous navigation systems. Autonomous navigation packages utilize visual odometry; Simultaneous Localization and Mapping (SLAM); and terrain contour matching to navigate between points when the UAV can no longer receive GPS position data. These systems use sensors such as high-definition cameras, LiDAR, and synthetic aperture radar (SAR) to provide near real-time geographical and obstacle data to the AI model. The AI model can then identify geographical landmarks and obstacles (e.g., mountains, buildings, and other aircraft) to ascertain its geographical location and autonomously navigate to its destination without the use of external satellite signals (Baykar Technology, 2026; Cybernews, 2026).

2.2 AI at the Targeting Level: Computer Vision and Multi-sensor Fusion

Target recognition is a complicated area when artificial intelligence (AI) is being used within military unmanned aerial vehicles (UAVs), from initial identifying of a target right through to the final elimination of that target. Deep learning algorithms such as CNNs (Convolutional Neural Networks) used in drones have been trained on large amounts of data, including images of military (military targets), civilian (civilian targets), and environmental (non-targets) objects (Donnelly & Wilson, 2026a). Algorithms process images using visual patterns, infrared (IR) signatures, and radar cross-sections to identify and classify an object in real time.

One of the significant technical issues involving AI and drones is related to environmental variables, adding complexity to the high density of modern battlefields. An Artificial Intelligence (AI) model must be able to accurately differentiate between an enemy tank that is concealed by trees and bushes from a civilian agricultural tractor parked almost exactly in the same spot (Donnelly & Wilson, 2026a). Modern drones are designed to achieve the required level of certainty necessary for lethal engagement in part due to their multi-sensor fusion systems. Multi-sensor fusion systems can process EO/IR (Electro-Optical/Infra-Red) data, acoustic data, and radar images allowing the AI to see under environmental conditions such as rain, fog, smoke, and when objects are cloaked with thermal masking. By processing the variety of input signals through sophisticated algorithms that produce a probability score for each potential target, the drone can establish when the total probability score for all potential targets that exceeds a predetermined algorithmic threshold. At this point, the drone's system locks onto the target's location and shortens the "last mile" from the kill chain. Therefore, when the munition is released, it will be able to autonomously guide a weapon toward the target without any human verification prior to impact, raising serious concerns about the accuracy of the AI's ability to identify target type (Cybernews, 2026; c).

2.3 Cyber Vulnerabilities: Adversarial Machine Learning and Sensor Spoofing

The increasing use of Deep Neural Networks (DNNs) carries with it a growing risk for Autonomous Drones that could fall victim to the new threat of Adversarial Machine Learning (AML). AML is a method of exploiting the inherent mathematical weaknesses of an ML model by causing small or even unnoticeable changes to input data, thus causing the AI to create faulty predictions or classifications (Xi et al. 2025). In the physical world, adversaries now use "adversarial patches", which are intentionally created visual patterns or codes (typically QR code-like), or which have altered textures, all placed on or near military assets. When an Autonomous Drone's electro-optical sensor captures an adversarial patch, the spatial displacement of the adversarial patch will disrupt the CNN's feature extraction process mathematically. This will result in the preventing the drone's object detection algorithm from identifying a main battle tank (making it invisible) or misidentifying the vehicle for a benign object (e.g. a rock formation/ civilian vehicle) (Xi et al. 2025).

In addition to being susceptible to complex physical-layer spoofs through AI-assisted drones, adversaries can use Generative Adversarial Networks (GANs) to approximate the actual probability distributions of GPS location and +communication signals. Attackers can then use delta (fake) signals to mislead UAV’s PID (proportional-integral-derivative) controller with incorrect spatial and temporal information. This input perturbation avoids traditional intrusion detection systems (IDS) and enables the ability of adversaries to reroute the drones’ course of flight, contend for the mission parameters set by the original drone operator, or create a controlled crash of the drone without the use of brute force or greater jamming bandwidth, which are both easily detectable (Erpek et al., 2024; Florida International University, 2024).

2.4 The Cyber-Kinetic Threshold

The development of autonomous drone swarms blurs the line between digital interference and physical, kinetic destruction in a way that changes our understanding of armed conflict between states. Cyber operations against the cognitive layers of drone swarms create an extreme ambiguity about how states should respond to such operations. One example would be if a state actor conducted a sophisticated cyberattack against an adversarial Targeting AI that subtly alters the Targeting weights or Confidence limits of an adversarial AI Drone Swarm, causing the adversarial drones to autonomously engage with civilian infrastructure as opposed to military infrastructure; in this case, the digital cyber operation could yield a catastrophic kinetic impact.

Under the traditional “Effects Based” framework of International Law, a state will only demonstrate a warlike Response to a Digital Cyberattack after the event has caused physical destruction equivalent to what would occur with an equivalent Kinetic Weapon employed by the state. This framework is thus outdated and insufficient for the digital age, and a Norms Based threshold should be the basis for measuring this new type of operation. In the Norms Based framework, the definition of a Cyberwarfare Act will be based on the type of Target attacked (e.g. a Target that is required for Human Life; a nuclear facility; a platform for Autonomous Weapons) or the intent of the Cyber Operation regardless of whether or not the Digital Attack caused the intended effects. This change will highlight that a single piece of malicious code embedded within an AI Navigation package of an AI supported Drone Swarm is no longer simply a tool for Intelligence Gathering; it is now engaging in an act of War.

3. Ethical Issues

The delegation of lethal authority to algorithmic systems introduces ethical friction, directly challenging the foundational principles of human agency, moral philosophy, and international law in warfare.

4. Case Studies

The following case studies illustrate the diverse, real-world applications of AI in contemporary unmanned systems, highlighting the technical realities, operational advantages, and ethical tensions of autonomous warfare across different geopolitical theaters.

4.1 Ukraine's Modular AI Kits and the "Last Mile" Kill Chain

The continuing military crisis in Ukraine has become an exceptional ‘real-time’ evaluation of the application of private sector (commercial) artificial intelligence (AI) and military AI technology integration. For example, having an enemy with significantly greater electronic warfare capacity than they possessed, Ukrainian forces determined that their typically low-cost, First-Person View (FPV) drones would be among the most vulnerable type of drones to radio frequency (RF) and GPS signal jamming. If the communication link between the FPV drone operator (often referred to as ‘pilot’ and the FPV drone was broken, that drone would likely become unusable, either crashing into a harmless area or not achieving the intended target location.

To address this problem, Ukraine has been using advanced modular AI kits (e.g., 33K Auterion Skynodes supplied by the US and Germany) on tens of thousands of FPV drones and loitering munitions in an iterative generation fashion supported by collaboration with their Western partners (Kirichenko, 2025; Military Tech Breakdown, 2025). The AI modules included in the FPV drones / loitering munitions serve as autonomous ‘brains’ of the vehicles, executing complicated computer vision algorithms on edge computing middleware eliminating the necessity for sustained cloud bandwidth to communicate back to a remote server for command-and-control purposes. The operator of the drone will manually fly it into the area near where they believe the target is located. Once the enemy target is seen on their display, the operator will draw a box around it. The AI will completely take over. The aircraft will no longer use GPS for location or video feeds to navigate and strike the locked target (Cybernews, 2026; Kirichenko, 2025; Military Tech Breakdown, 2025). This capability has changed the economics of war by enabling a lower-cost drone to destroy a high-value target such as an armored vehicle. In addition, Ukraine has developed a system called "Operational Gamification" to manage and exploit its use of strikes. By converting verified destruction across the battlefield into a digital currency (ePoints) through its DELTA situational awareness system and DOT-Chain logistics engine, tactical units can obtain additional drone capabilities (Debuglies, 2026). This system enables rapid engagement on the battlefield as the deployment of encrypted on-board AI software forces the elimination of human involvement in the decision-making process regarding lethal engagements during the final moments of the flight of the drone. The systemic risk of Goodhart's Law {State the Law] occurs as tactical inefficiencies and ethical violations can result from the deliberate removal of the human component from the kill chain due to the pursuit of "point targets" created by algorithms (CSIS, 2024; Debuglies, 2026).

4.2 Russia's KUB-BLA and Lancet Loitering Munitions

The KUB-BLA and the Lancet are the main loitering munitions used by the Russian Federation. Developed by the Zala Aero Group (a subsidiary of the Kalashnikov Concern), the KUB-BLA delivers a blast-fragmentation payload to specified target coordinates and has been claimed by its manufacturer to have high levels of autonomy. According to Kalashnikov, the KUB has Artificial Intelligence Visual Identification (AIVI) technology. This proposed system provides real-time target recognition and classification to potentially increase the area that could be covered by a single flight by a factor of 60 and improve the drone's level of autonomy and lethality (Automated Research, 2022).

The KUB-10M is a newer variant of this munition that uses the same electro-optical guidance system as the original KUB, but has an extended range of 100 kilometers and is fully reliant on electro-optical guidance rather than GPS, thus reducing its vulnerability to extensive GNSS jamming in the vicinity of the Ukrainian front lines (Cranny-Evans, 2026). However, there is a significant divergence between Russian promotional claims and the actual battlefield experience. The description "payload targeting image" appearing in KUB technical documentation describes a system that operates more like a traditional GPS guided munition, which would use fundamental contrast-locking methods to determine missile position, than a sophisticated deep-learning neural network (CSIS, 2022). KUB-BLA presents an ethical dilemma because of technological uncertainty and the policy implications of an AI arms race. If a state uses a primitive image contrasting software ‘disguised’ as highly advanced AI, the chance for misidentification increases exponentially.

The deployment of immature autonomous weapons that do not possess sophisticated contextual discernment capable of meeting the principles of distinction presents an extreme threat to civilian populations. (CSIS, 2022; Wilson & Donnelly, 2026a)

4.3 Turkey's Bayraktar Kamankeş 1 AI-Powered Cruise Missile

Baykar Tech is a major presence in the world of unmanned aerial warfare. From its original production of Bayraktar TB2 aircraft, it has developed and continues to develop new, intelligent, AI-enabled munitions. Kamankes 1 is an air-launched, jet-propelled cruise missile that is to be used against targets from larger flying launch platforms, including Akinci, TB2, and TB3 UAVs. The Kamankes 1 cruise missile represents an advance in using AI target acquisition systems leveraging state-of-the-art AI-enhanced EO (electro-optical) guidance systems to autonomously acquire and engage a target in flight (cruising at 140 knots / IDL) to a range greater than 100 km (Aviacionline, 2023; Baykar Technology, 2026). The Kamankeş 1 is more than a weapon designed to strike ground targets. In a recent test, the AI-guided Kamankeş 1 intercepted an airborne target during flight when fired from an Akinci unmanned combat aerial vehicle (UCAV) (Overell, 2025). This new air-to-air capability changes the framework that has defined drone warfare. It appears that AI-powered loitering munitions can be used as flexible, multi-domain platforms to autonomously conduct air defense missions. However, the high degree of autonomy this level of performance creates in three-dimensional airspace significantly complicates the role of human operators. An AI-powered, high-speed, jet-powered munition relies entirely on its AI-based visual identification system to assess the difference between an enemy drone and a friendly or civilian aircraft within crowded and contested airspace, pushing the conventional limits of algorithmic accuracy (Overell, 2025).

4.4 DARPA EVADE and Sikorsky MATRIX Software (USA)

The Defense Advanced Research Projects Agency (DARPA) is pushing to get autonomous systems into the field via the Early Vertical Takeoff and Landing Aircraft Demonstration (EVADE) program. EVADE is a part of DARPA’s Phase 2 acceleration of their ANCILLARY Program, focusing on the development of unmanned aerial systems in the Group 3 category (unmanned aerial vehicles weighing 330 pounds or less). EVADEs aim at designing unmanned aerial vehicles that function like a "flying truck." The vehicles are expected to use vertical take-off and landing (VTOL) capabilities in austere environments without requiring any dedicated infrastructure (DARPA, 2025). To reduce the time to the first flight and speed the deployment of the five contractors' designs, DARPA required all five EVADE contractor designs to utilize a single, standardized autonomy core: the Sikorsky MATRIX flight autonomy system. The MATRIX autonomous flight system was developed as part of DARPA's Aircrew Labor In-Cockpit Automation System (ALIAS) program and is viewed as the central nervous system for the contractor vehicles. The MATRIX flight autonomy system performs entire missions autonomously - all levels of flight control, complex navigation, and obstacle avoidance - from takeoff to landing, thereby significantly reducing the need for user input during long transit distances (DARPA, 2025; Easley, 2024).

The endurance requirement for the EVADE platforms is to be capable of operating for at least 12 hours over a range of 100 nautical miles, while carrying a payload of 60 pounds, performing various missions including logistics, weapons delivery, ISR and RSTA missions. The AVD drones will communicate directly with individual ground-based troops through the U.S. Navy's Battle Management System (BMS), eliminating the need for centralized ground control facilities entirely. This decentralization of command makes it possible to democratize airpower at the tactical level, and create large, distributed networks of autonomous agents. This was shown by the Army's most recent Swarm Forge exercise at Camp Blanding, where the U.S. was able to execute the first-ever kinetic drone swarm in history on U.S. soil (DefenseScoop. (2026). The use of AI to command-and-control multiple autonomous systems will require an extremely high level of cyber resilience. All distributed networks are at high risk for cyber-kinetic infiltration, adversarial AI attacks, and protocol manipulation; hence they require a very high level of cybersecurity.

5. Anticipatory Ethics

Addressing the ethical challenges posed by autonomous drones and cyber warfare requires robust, anticipatory ethical frameworks that guide the developers, software engineers, and military strategists designing these systems. Anticipatory Ethics is a framework which focuses on the identification of potential ethical issues relating to new and emerging technologies before the technologies become fully entrenched (Brey, 2012). The Association for Computing Machinery (ACM) Code of Ethics and Professional Conduct, combined with the

Anticipatory Technology Ethics (ATE) framework, provides a foundation for governing an Anticipatory Ethical analysis of military AI before it reaches the battlefield.

5.1 ACM Code of Ethics

Principle 1.1 (Contribute to humanity and society): This principle requires that computer professionals limit potential threats to the health, safety, and privacy of all people. Military AI developers for drones face a paradox. Designed to kill enemy combatants for state protection, they are obliged to limit any harms that may occur because of how their system was developed. For example, if unintended casualties result from asymmetric violence caused by their algorithms used in military operations (such as algorithm bias or misidentification), there will be no justification for these negative consequences (Association of Computing Machinery, 2018). Principle 1.2 (Do not do harm): The Code of Ethics provides guidance for work that is inherently designed to do harm to another person or entity but requires that "those responsible are obligated to provide ethical justification for the harm" (Association of Computing Machinery, 2018). There is also a requirement for professionals to conduct an in-depth analysis of the "emergent properties of the system". With respect to autonomous drone operations (swarms), professionals will need to model, simulate, and mitigate any unintended escalation of violence that results from the emergent behaviors of multiple drones interacting with one another in ways that were not intended and thus could not be anticipated before deploying the swarm of drones.

5.2 Anticipatory Ethics

Before mass production and deployment occur, principles found within the Anticipatory Technology Ethics (ATE) framework developed by Philip Brey (2012) should be applied to enforce ethical considerations regarding the deployment of lethal autonomous weapon systems (AWS). In many cases, traditional ethics only evaluates harmful impacts of technology after technology has been deployed. ATE approaches ethical assessment from an anticipatory perspective, focusing on identifying and assessing ethical concerns at both the research & development (R&D) and before the deployment stages of new technology (Brey, 2012). For ATE to work as intended, methodologically sound foresight and future studies must occur throughout the development of emerging technologies, particularly because there is uncertainty about the development and deployment of many emerging technologies. The assessment of new technologies must include continuous ethical assessment considering three levels within the ATE framework: first, at the technology level (e.g., bias found in computer vision); second, at the artifact level (e.g., evaluating the specific technology of the Kamankeş 1 loitering munition); and third, at the application level (e.g., modeling the potential ethical impacts of using the Kamankeş 1 loitering munition in an urban combat environment) (Brey, 2012). When engineers apply a "Sociotechnical Imperative" to drone development employing ATE, they must understand that self-flying drones are not functioning in a vacuum. Drones are integrated into politically unstable and dynamic environments. Developers must also feel a moral obligation (pursuant to the "Foreseeability of Effect") to identify and develop strategies to mitigate any potential exploitation of their new systems through either hacking, spoofing, reverse engineering, or influence through multifactional information. Engineering an autonomous weapon that lacks a safe mechanism (e.g., a reliable recall, cryptographic verification, or failsafe behaviors), violates the sociotechnical imperative because it is entirely possible that the weapon could be manipulated through adversarial machine learning to kill innocent civilians (Association for Computing Machinery, 2018; Wilson & Donnelly, 2026b).

6. Recommendations

To resolve the technical vulnerabilities and ethical issues identified in the case studies, military developers and policymakers must move beyond abstract principles and develop, mathematically verifiable engineering guardrails. The following technical recommendations bridge the critical gap between anticipatory ethics and operational cyber-resilience.

6.1 Adversarial Training and Data Diodes for Sensor Integrity

Developers must approach the redesign of drone Intrusion Detection Systems (IDSs) and protect against adversarial machine learning (AML) attacks to protect autonomous systems such as the KUB-BLA or Kamankeş. 1. To accomplish this, developers must implement rigorous adversarial training frameworks using Generative Adversarial Networks (GANs). This adversarial training will occur during training by continuously sending highly sophisticated GAN-generated adversarial patches and simulated global-positioning-system (GPS) spoofing signals (delta) into the neural networks of the drones to teach the drones to recognize and reject any input perturbation from normal probability distributions. By using this dynamic, the drone isn't static and is continually improved to identify the mathematical signatures of cyber deception, thus improving the detection rate of cyber deception across all three primary areas of focus in IDSs, namely physical layer security, detection rate accuracy, and detection rate of cyber deception (Erpek et al., 2024; Florida International University; 2024; Xi et al., 2025). In addition, flight control subsystems are logically and physically separated from any external data links using Data Diodes, which are hardware devices that allow information to only flow in one direction. Essentially, the drone can transmit telemetry, camera feeds and situational awareness information to the Battle Management System (such as what is being done with the DARPA EVADE program). However, the flight controller cannot receive any unverified external command data from the internet, preventing an adversary from using the drone's C2 link to change its flight path or release its payload (Wilson& Donnelly, 2026a).

6.2 Algorithmic Circuit Breakers and Hard-Coded "Ethical Autonomy"

To ensure safety and security during war and cyber warfare, there is a need for a reliable fail-safe system to regulate machine-led weapons systems. One way to accomplish this is to require all military weapon systems that are automated to have an independent means of verification called "Algorithmic Circuit Breakers." These breakers should function independently from the Primary Targeting AI and monitor the drone's operating parameters and confidence intervals at regular intervals to maintain control over land or airspace.

If a drone has lost communication with its operator (such as an autonomous Ukrainian first-person view drone operated from a Skynode module) and then subsequently experiences a drop in its confidence interval for identifying a target (e.g., as indicated by a score below 99.9%), the circuit breaker will be automatically invoked. Additionally, a predetermined series of failsafe conditions that are ethically based should also be hard coded into the system to trigger abort criteria. For instance, the system should be programmed to either (a) remain in place until communications are established, (b) self-destruct within an established hygienic area, or (c) return to base. (Kugler, 2026). The targeting system cannot be allowed to adjust its confidence level downwards to allow a strike due to reduced data availability. This system would directly limit the likelihood of violations of LOAC and allow for the safe failure of the machine to preserve human life in cases of uncertainty (Kugler, 2026). Additionally, parameters for cyber-kinetic escalation must be established for swarm operations. Once a swarm discovers that it is conducting a cyber battle against an adversarial AI defensive system at machine speed, a circuit breaker will halt any offensive action once a predetermined "heat level" (i.e. a specific volume of cyberattacks in a specific period) has been reached. This endpoint forces a stoppage for human assessment and prevents a localized digital conflict from escalating unintentionally into a full-scale "Flash War" (Wilson & Donnelly, 2026a).

7. Future Work

Due to this rapid and disproportionately increasing trend involving drones, there is an urgent need for more extensive and ongoing research into future generations of autonomous threats from militarized artificial intelligence. Future research will be required to assess both the performance of autonomous munitions individually and the various interactive dynamics involved within a fully autonomous combat environment. "Agentic AIs" could autonomously negotiate and plan for and perform multiple simultaneous operations in cyberspace and via kinetic means across all environments (land, sea, air, and airspace) without human intervention, prior programming of waypoints, or definitive sequential steps. There is a need to investigate the moral dimension of these algorithms: whether an Agentic AIs can be coded to understand and apply the nuances of the law of armed conflict (LOAC), and whether the absence of human empathy in the overall strategic planning process will inherently lead to the perpetration of atrocities.

8. Conclusion

The integration of AI, next generation drone technologies and advanced cyber warfare capabilities is defining an entirely new generation of warfare with never-before-seen speed, volume, independence (autonomy) and lethality. The tactical advantage created by developing systems to complete the "last mile" of the kill chain without relying on vulnerable telecommunications, as demonstrated by current systems in Ukraine (Skynode-equipped drones), Turkey (Kamankeş cruise missiles) and the US (DARPA EVADE), removes the ability for traditional electronic warfare countermeasures to defend against them, as military tactics are merging with these commercial aerospace enablement's utilizing on-board (edge) computing to allow the waning effectiveness of traditional EW countermeasures (Baykar Technology, 2026) (DARPA, 2025) (Wilson & Donnelly, 2026a) (Military Tech Breakdown, 2025).

As a result, the change in how wars are fought (from traditional attrition warfare to algorithmic warfare) introduces ethical/legal/strategic costs. The transfer of deadly decision-making authority to machines creates a challenge to the laws of armed conflict and erodes the core ethical principles of distinction and proportionality; it also creates a virtually insurmountable accountability "black hole" where war crimes may never be punished (i.e., software glitches). Additionally, since these lethal machines utilize deep learning neural networks, they are vulnerable to psychological cyber vulnerabilities like adversarial machine-learning attacks, sensor spoofing, and protocol manipulation, enabling adversarial forces to disrupt a drone's perception of the world by using adversarial techniques to control the drone's perception and turn that country's weapons against its own citizens/civilian infrastructure (Santoni de Sio & van den Hoven 2018; Schwarz 2024; Erpek et al. 2024; Xi et al. 2025).

For human beings to maintain their moral agency and stable strategy in times of machine-enabled speed warfare, developers, commanders of military forces, and world leaders will need to change how they think about AI, no longer regarding it as just "another piece of equipment in their arsenal," but rather, recognizing it as a independent and probabilistically based agent requiring the imposition of strict, globally agreed upon and Ethically Based Controls on its use. Those using AI and all those affected by its use must have a commitment to following the guidelines of Anticipatory Technology Ethical Guidelines (ATE) and ACM Code of Ethics; doing so becomes a prerequisite for humanity's creating and retaining the ability to operate with safety (Association for Computing Machinery 2018; Brey 2012). The mandatory inclusion of adversarial training methods, multi-sensor fusion designs, documented provenance for all data sources used, and hard-wiring algorithmic circuit breakers into all AI systems can no longer be seen as just "technical preferences;" they must be ethically required. Without these mathematically verifiable, robust countermeasures, the speed at which autonomous warfare occurs will outpace our ability to restrain ourselves, thus converting tactical advantage into systemic, unmanageable lethality across global borders..

Ethics Declaration: No human participants or personally identifiable information were involved. All data sources were publicly available.

AI Tools Declaration: ChatGPT 5.1 for drafting and refinement. Human authors verified all content. Gemini 3.0 pro used for aiding in sourcing.

References

Association for Computing Machinery. (2018). ACM Code of Ethics and Professional Conduct. https://www.acm.org/code-of-ethics

Automated Research. (2022). Weapons Systems with Autonomous Functions Used in Ukraine. https://automatedresearch.org/news/weapons-systems-with-autonomous-functions-used-in-ukraine/

Aviacionline. (2023). Baykar unveils new loitering ammunition for drones with more than 50 km range. https://www.aviacionline.com/baykar-unveils-new-loitering-ammunition-for-drones-with-more-than-50-km-range

Baykar Technology. (2026). KEMANKEŞ 1 AI-Powered Mini Cruise Missile Technical Specifications. https://baykartech.com/en/uav/kemankes/

Brey, P. A. E. (2012). Anticipatory Ethics for Emerging Technologies. NanoEthics, 6(1), 1-13. https://doi.org/10.1007/s11569-012-0141-7

Center for Strategic and International Studies. (2022). Russia Probably Has Not Used AI-Enabled Weapons in Ukraine, but That Could Change. https://www.csis.org/analysis/russia-probably-has-not-used-ai-enabled-weapons-ukraine-could-change

Center for Strategic and International Studies. (2024). Ukraine's Future Vision and Current Capabilities: Waging AI-Enabled Autonomous Warfare. https://www.csis.org/analysis/ukraines-future-vision-and-current-capabilities-waging-ai-enabled-autonomous-warfare

Communications of the ACM. (2024). Multi-Use AI and Moral Responsibility. https://cacm.acm.org/opinion/multi-use-ai-and-moral-responsibility/

Cranny-Evans, S. (2026). Kub-10ME: Kalashnikov launches new loitering munition with 100 km range. Calibre Defence. https://www.calibredefence.co.uk/kub-10me-kalashnikov-launches-new-loitering-munition-with-100-km-range/

Cybernews. (2026). AI-powered military drones and the future of warfare. https://cybernews.com/ai-news/ai-powered-military-drones-future-warfare/

DARPA. (2025). DARPA to demonstrate revolutionary drone capabilities for warfighters. https://www.darpa.mil/news/2025/evade-drone-capabilities-warfighters

Debuglies. (2026). Beyond Drones: How DELTA & DOT-Chain Created the First AI Battlespace. https://debuglies.com/2026/01/07/beyond-drones-how-delta-dot-chain-created-the-first-ai-battlespace/

DefenseScoop. (2026). Drone Swarm Forge Demonstration at U.S. Military Camp Blanding. https://defensescoop.com/2026/01/15/drone-swarm-forge-demonstration-us-military-camp-blanding/

Easley, M. (2024). DARPA to test autonomous flight capability on Army's Black Hawk helicopter in 2025 . DefenseScoop. https://defensescoop.com/2024/10/14/darpa-sikorsky-black-hawk-autonomous-flight-award/

Erpek, T., et al. (2024). Detection of GPS Spoofing Attacks in UAVs Based on Adversarial Machine Learning Model . Sensors, 24(18), 6156. https://www.mdpi.com/1424-8220/24/18/6156

Florida International University. (2024). Adversarial Data Augmented Resilient IDS for UAVs. https://acyd.fiu.edu/wp-content/uploads/Adversarial_Data_Augmented_Resilient_IDS_for_UAVs__PSBD.pdf

Kirichenko, D. (2025). Ukraine's AI Drones Hunt the Enemy. Center for European Policy Analysis. https://cepa.org/article/ukraines-ai-drones-hunt-the-enemy/

Kugler, L. (2026). From Drones to Deepfakes: How AI is Reshaping War. Communications of the ACM. https://cacm.acm.org/news/from-drones-to-deepfakes-how-ai-is-reshaping-war/

Overell, J. (2025). Baykar Demonstrates Air-to-Air Capability with Kemankes 1. SAE Media Group. https://www.smgconferences.com/editors-corner/6035-news--baykar-demonstrates-airtoair-capability-with-kemankes-1-loitering-munition

Santoni de Sio, F., & van den Hoven, J. (2018). Meaningful Human Control over Autonomous Systems: A Philosophical Account. Frontiers in Robotics and AI, 5, 15. https://doi.org/10.3389/frobt.2018.00015

Schwarz, E. (2024). The Ethical Implications of AI in Warfare. Queen Mary University of London. https://www.qmul.ac.uk/research/featured-research/the-ethical-implications-of-ai-in-warfare/

Stop Killer Robots. (2022). Increasing Autonomy in Weapons Systems. https://www.stopkillerrobots.org/wp-content/uploads/2022/10/Report-Increasing-Autonomy-in-Weapons-Systems-Single-page-viewfp.pdf

Wilson, R., & Donnelly, N. (2026a). Cyber Operations and the Threshold for Cyber Warfare: Ethical and Anticipated Ethical Issues. Proceedings of the 21st International Conference on Cyber Warfare and Security (ICCWS 2026), 578-584. https://papers.academic-conferences.org/index.php/iccws/article/view/4504

Xi, H., Ru, L., Jiwei, T., & Luan, X. (2025). URAdv: A Novel Framework for Generating Ultra-Robust Adversarial Patches Against UAV Object Detection. ResearchGate. https://www.researchgate.net/figure/Proposed-adversarial-patch-generation-framework_fig2_388947904