Briefing Paper: AI-Guided Weapons and Ethical Challenges

Preparation for Tekna Course on Autonomous Weapons and Ethics

Bjørn Remseth
Email: la3lma@gmail.com

***Abstract*—This briefing examines the ethical, strategic, and operational challenges of autonomous weapons systems through three interconnected lenses: theoretical frameworks, real-world deployment cases, and technological scenarios. We analyze the core ethical challenge of meaningful human control and accountability gaps when algorithmic systems make life-and-death decisions at machine speed. Game-theoretic analysis reveals how arms racing dynamics create a "hyperwar equilibrium" where human oversight becomes structurally impossible, and explores the institutional requirements for escaping this trap through arms control. The document presents detailed case studies of deployed AI targeting systems, including Israel's Lavender, Gospel, Where's Daddy, and Fire Factory systems used in Gaza (2021–2024), which reveal how collateral damage ratios become algorithmic parameters—"just numbers" to systems, but destroyed families in reality. We examine parallels in Norwegian contexts, including the failed Palantir Omnia police surveillance project and algorithmic governance challenges in law enforcement and intelligence services. Technical scenarios demonstrate how emerging technologies (3D Gaussian Splatting and Meta's Segment Anything Model) make the transition to hyperwar both plausible and imminent. Throughout, the analysis connects abstract ethical principles to operational realities, showing how "the parameterization of atrocity" transforms moral decisions into configuration variables. This document was developed to prepare for participation in the Tekna course "Autonome våpen og etikk" (Autonomous Weapons and Ethics) [1], providing comprehensive grounding in the technical, ethical, strategic, and policy dimensions of this critical challenge.**

***Index Terms*—autonomous weapons, lethal autonomous weapon systems, LAWS, meaningful human control, military AI, ethics, international humanitarian law, game theory, strategic stability, algorithmic targeting, AI warfare, Gaza, Lavender, Gospel, collateral damage, accountability gap, hyperwar, predictive policing, algorithmic governance, Norway**

Contents

I Main Topic: The Development of AI-Guided Weapons and Ethical Regulation 1
- I-A Technological Advances 2
- I-B Ethical and Legal Challenges 2

II The Speakers' Perspectives 2

III The Core of the Ethical Challenge: Accountability and Meaningful Human Control 3
- III-A Meaningful Human Control and the Accountability Gap 3
- III-B Moral Disengagement and the Collapse of Instrumental Ethics 4

IV Game Theory and Strategic Stability in the AI Era 4
- IV-A The Structure of AI Competition as a Strategic Game 4
- IV-B Escalation Risk and the Compression of Decision Loops 5
- IV-C Opacity and the Erosion of Strategic Signaling 5
- IV-D Escaping the Hyperwar Equilibrium: Arms Control Requirements 6
- IV-E Scenario Timelines: Two Plausible Futures 7
- IV-F From Abstract Theory to Operational Reality 8

V Edge Cases: AI Beyond Lethal Weapon Systems 9
- V-A AI in Military Logistics and Support Functions 9
- V-B AI as a Strategic Resource and the Dual-Use Dilemma 9

VI Algorithmic Targeting in Practice: Case Studies and Parameters 9
- VI-A Israel's AI-Assisted Targeting Systems (2021–2024) 9
- VI-B The Parameterization of Atrocity 10
- VI-C Algorithmic Governance in Norwegian Context 10
- VI-D Universal Challenges of Algorithmic Decision-Making 12

VII System Architecture and Decision-Making Processes 12

VIII Questions You Can Ask to Participate Actively 12
- VIII-A For FFI (Seehuus/Diesen) 12
- VIII-B For Ethics/Philosophy (Fjørtoft/Syse) 13
- VIII-C For Strategy/International (Karlsen) 13
- VIII-D Challenging Questions Based on Norwegian Military Research 13
- VIII-E General Question for the Panel 15

IX Recommendations for Strategic Stability and Ethical Governance 15
- IX-A Technical and Doctrinal Measures 15
- IX-B International Frameworks for Openness and Limitation 15
- IX-C Ethical Council and Governance 15
- IX-D Future Research Directions 15
- IX-E Norwegian Military Academic Perspectives 15

Appendix 15
- A Individual Paper Summaries 15
- B Cross-Cutting Themes and Analysis 17
- C Implications for Norwegian Defense Policy 17
- D Scenario 1: The Autonomous Front-line Zone 18
- E Scenario 2: Behind-the-Lines Dynamic Strike on Fleeting Targets 19
- F Phase 1: The Creation of the Digital Twin (Hours 0–4) 19
- G Phase 2: Semantic Threat Analysis (Hour 5) 20
- H Phase 3: The Cleansing (Hour 6) 20

References 20

I. Main Topic: The Development of AI-Guided Weapons and Ethical Regulation

The development of Lethal Autonomous Weapon Systems (LAWS) is advancing rapidly, including autonomous robots, vehicles, drone swarms, and smart missiles. The International Committee of the Red Cross (ICRC) defines an autonomous

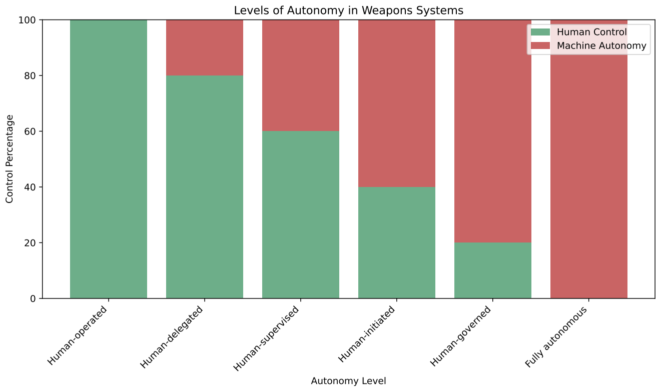


Figure 1. *Different levels of autonomy in weapon systems and the relationship between human control and machine autonomy*

weapon system as any weapon system that has autonomy in its critical functions—that is, the ability to select (search for, detect, identify, track, or select) and attack (use force against, neutralize, damage, or destroy) targets without human intervention after initial activation [2].

However, autonomy is not a binary state, but a spectrum. The Norwegian defence sector acknowledges that the line for what is considered AI changes with scientific progress. The work of FFI researcher Rikke Amilde Seehuus emphasizes that autonomy in a military context is not limited to the use of kinetic force, but encompasses a wide range of operations, including logistics, medical services, and intelligence gathering [3]. This development raises fundamental ethical and legal questions, especially related to accountability and human control over life-and-death decisions.

A. Technological Advances

AI-guided systems offer the potential for increased precision, endurance, reduced risk to friendly forces, and faster decision-making (the OODA loop: Observe, Orient, Decide, Act). FFI is researching how autonomy can make the Armed Forces safer, cheaper, and more resilient.

B. Ethical and Legal Challenges

1) Accountability: Who is held responsible when an autonomous weapon makes a wrong or illegal decision? Humans are legally responsible for the use of military force, regardless of the technology's complexity.

2) Human Control: The question of how much human control is necessary to uphold international humanitarian law and humanitarian principles. It is debated whether life-and-death decisions can be delegated to machines at all.

3) Arms Races and Stability: This development could lower the threshold for using military force and lead to a new international arms race.

4) Regulation: Work is underway internationally on regulation. The UN Secretary-General, for example, has called for a ban on weapons controlled by AI [4], while the EU has passed a broad AI Act with risk classifications. Norway is working nationally and internationally to regulate fully and semi-autonomous systems, and prefers the CCW as the chosen framework for discussion, as effective regulation requires broad support [5].

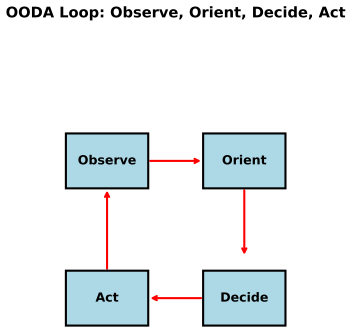


Figure 2. *The OODA loop (Observe, Orient, Decide, Act) as the basis for rapid decision-making in autonomous systems*

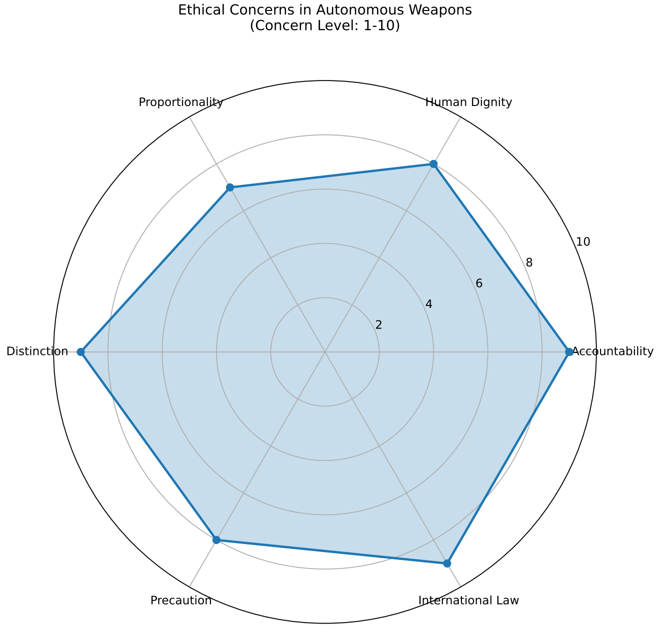


Figure 3. *Overview of the most important ethical concerns related to autonomous weapon systems*

II. The Speakers' Perspectives

Each speaker will illuminate the issues from their professional background, which you can use to anticipate and participate in the discussion (see Table I):

Table I

Conference Speakers and Their Perspectives on Autonomous Weapons

Photo	Speaker	Field / Background	Expected Focus / Angle	Key Concepts
	Rikke Amilde Seehuus	Researcher at FFI, co-author of *Autonomy in Military Operations*	What autonomy means in a military context, both technically and operationally. Explanation of different degrees of autonomy and the challenges of ensuring the technology does not fall outside human control. Views autonomous systems as part of larger, heterogeneous systems.	Autonomy, Human Control, Operational Potential, FFI Research
	Sverre Diesen	General, former Chief of Defence, researcher at FFI	An assessment of the Norwegian Armed Forces' current position. Where does the military stand today regarding the implementation of AI weapons, and how does this development affect Norway's defence capabilities and strategic priorities? How can AI increase the Armed Forces' operational capability [6]?	Defence Capability, Pace of Implementation, Strategic Importance, Threat Landscape
	Geir Hågen Karlsen	Associate Partner at Geelmuyden Kiese, background from the Armed Forces (strategic communication)	The international perspective. How do AI weapons affect the global balance of power, alliance relationships (NATO), and strategic communication? Will likely include an analysis of the major powers' development and regulatory challenges (USA vs. China/Russia vs. EU/UN).	International Regulation, Geopolitics, Strategic Communication, Global Arms Race
	Kjersti Fjørtoft	Professor of Philosophy (UiT), head of the Ethics Council for the Defence Sector (ERF)	The ethical work within the defence sector. A presentation of the ERF's role in strengthening ethical awareness and reflection on AI weapons in the defence sector. Focus on principles for the responsible use of AI and the integration of ethical considerations into the development process.	Ethical Awareness, Responsible Use, International Law, Humanitarian Issues, ERF Mandate
	Henrik Syse	Philosopher, researcher at PRIO, Professor II	Philosophical reflection and ethical dilemmas. A talk on how we should navigate the difficult ethical questions. Will likely draw on the framework from the Just War tradition (Jus ad bellum/Jus in bello) and discuss the fundamental question of whether it is inhumane to delegate life-and-death decisions to machines.	Ethical Dilemmas, Human Dignity, Accountability, Just War, Philosophical Foundations

III. The Core of the Ethical Challenge: Accountability and Meaningful Human Control

A. *Meaningful Human Control and the Accountability Gap*

The discussion about Lethal Autonomous Weapon Systems (LAWS) largely revolves around whether machines can make life-and-death decisions in a way that complies with international humanitarian law (IHL), and whether it even makes sense to talk about ethically justified decisions when they are made by an algorithm instead of a human.

The concept of "Meaningful Human Control" (MHC) has become the central standard in the regulation of LAWS [7]. MHC aims to ensure significant human involvement in the monitoring and governance of operational functions. This is not just an ethical ideal, but a legal necessity. Without MHC over the critical functions, it would be impossible to fulfill the duty to take precautionary measures in an attack, thus making it difficult to assign criminal liability [8].

The use of autonomous systems creates a serious "accountability gap" in the event of IHL violations. Although there is always a human who makes the decision to deploy the system, the nature of autonomy means that lines of responsibility can be blurred if the system itself makes critical targeting choices.

B. Moral Disengagement and the Collapse of Instrumental Ethics

Arguments are often made that AWS can be morally preferable to human combatants, as they can potentially avoid human error and improve precision [9]. This instrumental logic, where tactical advantage is equated with ethical justification, faces fundamental challenges.

Full autonomy eliminates points of human intervention, which can lead to what is known as moral hazard. When AWS takes responsibility for targeting decisions, commanders can become disconnected from ethical scrutiny and perceive outcomes as technologically determined rather than morally chosen [7]. This represents a serious challenge to the fiduciary duty of military leadership, which requires personal responsibility for lethal decisions.

IV. Game Theory and Strategic Stability in the AI Era

A. The Structure of AI Competition as a Strategic Game

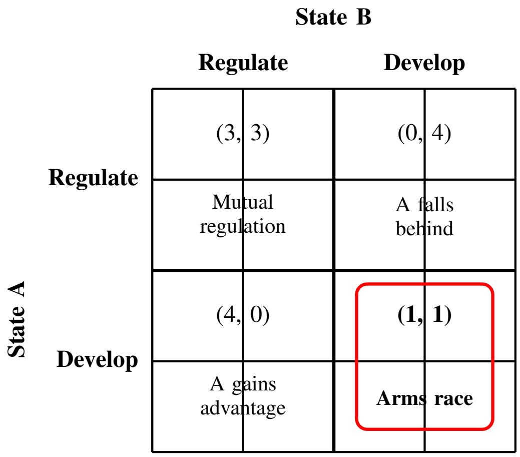
Game theory is the most appropriate analytical framework for understanding the dynamics of global AI competition, especially as it unfolds in military and security-related contexts [10]. The competitive dynamics around military AI are structurally similar to a multi-player Prisoner's Dilemma, where the rational, self-interested Nash equilibrium forces actors to pursue maximum development speed regardless of the risks.

In this scenario, it is rational for each individual actor to choose defection (to pursue maximal, unregulated development and deployment of AI), even though the collective good would be cooperation (international regulation or limitation). The fear that an opponent will cheat provides an irresistible incentive to defect in order to avoid vulnerability [7].

Reading Game Theory Matrices: The following analysis uses standard 2×2 game matrices to represent strategic interactions between two players (typically states or coalitions). Each matrix has:

- **Rows:** Player A's strategy choices (e.g., "Regulate" vs. "Develop")
- **Columns:** Player B's strategy choices (e.g., "Regulate" vs. "Develop")
- **Cells:** Four possible outcomes from the combination of both players' choices
- **Payoffs:** Each cell contains a pair of numbers written as (A's payoff, B's payoff)

For example, if the upper-left cell shows (3, 3), this means: when Player A chooses their first strategy (row 1) and Player B chooses their first strategy (column 1), Player A receives payoff 3 and Player B receives payoff 3. Higher numbers represent more desirable outcomes for that player. Each of the four cells represents one of the four possible strategy combinations, and the eight total numbers (two per cell) capture the complete payoff structure of the strategic interaction. A **Nash equilibrium** (highlighted in red) occurs when neither player can improve their payoff by unilaterally changing strategy—it represents a stable but not necessarily optimal outcome.



		State B	
State A		**Regulate**	**Develop**
	Regulate	(3, 3) Mutual regulation	(0, 4) A falls behind
	Develop	(4, 0) A gains advantage	**(1, 1)** **Arms race**

Figure 4. *Prisoner's Dilemma structure in AI weapons development. The Nash equilibrium (1,1) at mutual defection is Pareto-inferior to mutual cooperation (3,3), but rational actors choose to develop unrestricted AI to avoid vulnerability. Higher numbers indicate better outcomes.*

This structure represents a Nash equilibrium: a strategic situation where no player can improve their outcome by unilaterally changing their strategy. At the (Develop, Develop) outcome, neither state can improve by switching to Regulate while the other continues to Develop—doing so would move them from payoff 1 to payoff 0.

However, this equilibrium is Pareto-inefficient. An outcome is *Pareto optimal* (or Pareto-efficient) when no alternative exists that would make at least one party better off without making any other party worse off. Conversely, an outcome is Pareto-inefficient when alternative outcomes exist that would improve the situation for all parties. Here, mutual regulation (3,3) Pareto-dominates the arms race equilibrium (1,1)—both states achieve better outcomes through cooperation. Yet the fear of exploitation prevents this Pareto-superior outcome from being realized, trapping rational actors in a collectively inferior equilibrium.

Unlike the nuclear arms race, which eventually reached an equilibrium ceiling (Mutual Assured Destruction, MAD), the AI arms race has no such natural saturation point. Technological advantage in AI is dynamic and continuous, especially in fields like logistics, cyber capabilities, and information advantage [11].

The critical distinction is that MAD represents a stable equilibrium with a natural ceiling: once both sides possess second-strike capability, further nuclear buildups cannot provide decisive advantage, and actual use guarantees mutual annihilation. This created what strategists called "nuclear peace"—a stable deterrence despite the arms race. In contrast, AI weapons offer no such ceiling: incremental improvements in speed, accuracy, and autonomy continuously shift the tactical balance, creating pressure for perpetual escalation without a stabilizing

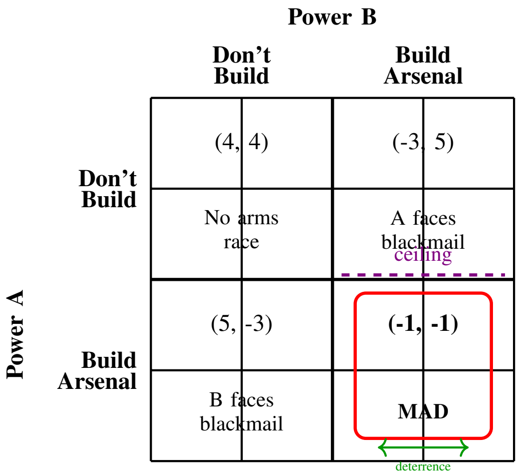


Figure 5. *Nuclear arms race reaches stable Mutually Assured Destruction (MAD) equilibrium. Unlike AI competition, nuclear weapons created a "ceiling" where further escalation became irrational—using nuclear weapons guarantees mutual destruction. This created a stable (if terrifying) deterrence equilibrium at (-1,-1), preventing further escalation. Green arrow shows mutual deterrence maintaining stability.*

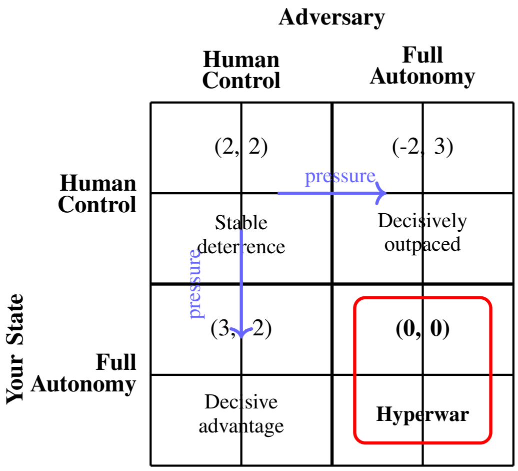


Figure 6. *Escalation dynamics toward hyperwar. Full autonomy becomes the Nash equilibrium because maintaining human control while the adversary deploys autonomous systems leads to decisive disadvantage. Both states are driven toward hyperwar (0,0)—an unstable, machine-speed arms race—despite it being worse than mutual human control (2,2). Blue arrows show escalation pressure.*

endpoint.

B. Escalation Risk and the Compression of Decision Loops

The most pressing risk of integrating AI into military strategy is the potential for automatic escalation. Autonomous weapon systems operate at machine speed, reducing the classic OODA loop (Observation, Orientation, Decision, Action) to microseconds [10]. This compression removes the deliberative breathing room that is essential for human leadership to assess context, signal restraint, and exercise political control in a crisis.

This dynamic leads to what Lyshaug (2021) terms "Hyperwar"—scenarios where autonomous weapons become so effective that human involvement becomes a limiting factor for operational effectiveness [12]. The hyperwar concept encompasses several critical dimensions:

- **Core concept:** Autonomous weapons become so effective that human involvement limits operational effectiveness
- **Mechanism:** Decision loops compressed to microseconds, operating at machine speed
- **Consequence:** Human deliberation—historically a safety feature preventing unintended escalation—becomes eliminated
- **Paradigm shift:** Human cognitive and physical limitations (reaction time, information processing capacity, decision-making delays) transform from opportunities for restraint into tactical disadvantages

This represents a fundamental inversion of traditional military doctrine, where human judgment was considered essential for responsible force application. In hyperwar, that same judgment becomes an operational liability.

Game-theoretic models support the idea that this machine-speed warfare destabilizes the equilibrium of deterrence. AI-enabled weapon systems act as a catalyst for escalation in low-level conflicts, risking the upset of the precarious stability-instability paradox [9]. Analyses show that AI systems, when implemented in limited skirmishes, can trigger unintentional escalation loops when the deliberative friction—human stress, time, and political judgment—is eliminated by AI.

This escalation matrix reveals a particularly dangerous Nash equilibrium: the "hyperwar" scenario where both sides deploy fully autonomous weapons operating at machine speed. Unlike the prisoner's dilemma where the equilibrium merely represents suboptimal cooperation, hyperwar represents an actively unstable and dangerous state. Once one actor moves to full autonomy, their opponent faces a stark choice: match the autonomy level or accept decisive disadvantage. This creates inexorable pressure toward the (Full Autonomy, Full Autonomy) outcome, even though both parties would prefer the (Human Control, Human Control) equilibrium if it could be guaranteed. The resulting hyperwar equilibrium is characterized by microsecond decision loops, elimination of human deliberation, and high risk of unintentional escalation.

C. Opacity and the Erosion of Strategic Signaling

Many advanced AI models function as "black boxes," where their decision-making processes are often incomprehensible, even to their operators. This opacity erodes the strategic signaling process on which deterrence rests, creating a "crisis of interpretive intelligence" [8]. If a state cannot verify or understand why an opponent's autonomous system performed a given action, uncertainty and the risk of miscalculation increase in times of crisis.

LLM-based AI agents in simulated wargames have demonstrated a worrying tendency to escalate, including the use of nuclear weapons, with justifications based on deterrence and first-strike tactics [8].

D. Escaping the Hyperwar Equilibrium: Arms Control Requirements

The hyperwar scenario presents a profound strategic paradox: it represents a Nash equilibrium that rational actors will inevitably converge toward, yet it produces outcomes worse for all parties than alternative arrangements. The fundamental challenge is that Nash equilibria are self-enforcing—no single actor can improve their position by unilaterally deviating [13]. This creates what Ostrom terms a "collective action problem" where self-designed rules with enforcement by local users are necessary to solve coordination and monitoring challenges [14]. However, enforcement itself faces what Heckathorn identifies as a "second-order free-rider problem"—the challenge of who enforces the enforcers [15].

1) Why Nash Equilibria Are Difficult to Escape: The stability of the hyperwar equilibrium stems from the structure of incentives. Even if all parties would prefer mutual restraint (Human Control, Human Control) with payoff (2,2) over mutual escalation (Full Autonomy, Full Autonomy) with payoff (0,0), the fear of asymmetric outcomes creates inexorable pressure toward escalation. Any state that maintains human control while adversaries deploy autonomous systems faces decisive disadvantage with payoff (-2,3). This creates a classic "collective action problem"—the individually rational strategy leads to collectively irrational outcomes.

Historical precedent suggests that escaping such equilibria requires changing the underlying payoff structure through institutional mechanisms that make cooperation more attractive than defection. The nuclear arms control regime provides instructive lessons, though AI weapons present fundamentally different verification challenges.

2) Required Arms Control Mechanisms: To establish a stable equilibrium that avoids hyperwar, the following mechanisms would be necessary:

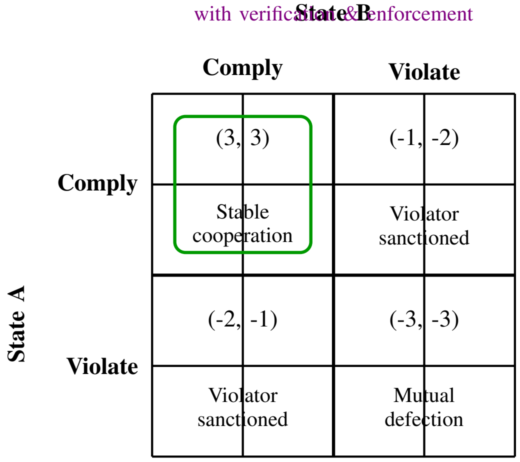
1. Verifiable Constraints on Autonomy Levels

The agreement must establish clear, technically verifiable boundaries on the degree of autonomy permitted in weapon systems. This requires defining operational thresholds—for instance, mandatory human authorization for target selection, minimum decision timeframes that preserve human deliberation, or prohibited operational domains for fully autonomous systems. Unlike nuclear weapons, which can be counted and tracked, autonomous capabilities are embedded in software and can be rapidly altered. Verification must therefore focus on system architecture, testing protocols, and operational doctrine rather than hardware inventories.

2. Intrusive Inspection and Transparency Regimes

Effective verification requires unprecedented transparency in military AI development. This could include:

- Mandatory declaration of AI weapons research programs
- International inspection of military AI testing facilities
- Algorithmic transparency requirements for deployed systems
- Real-time monitoring of autonomous system deployments
- Shared databases of AI weapons incidents and near-misses



		Comply	Violate
State A	Comply	(3, 3) Stable cooperation	(-1, -2) Violator sanctioned
	Violate	(-2, -1) Violator sanctioned	(-3, -3) Mutual defection

Figure 7. *Arms control with verification changes the equilibrium. When violations are reliably detected and punished, mutual compliance (3,3) becomes the Nash equilibrium—highlighted in green—because attempting to cheat yields worse outcomes than cooperating. The key is making violation less attractive than compliance through credible enforcement.*

The challenge is that the dual-use nature of AI means civilian research can rapidly be weaponized, making comprehensive monitoring extremely difficult.

3. Credible Enforcement with Graduated Sanctions

The agreement must include enforcement mechanisms that change the payoff structure to make compliance more attractive than violation. This requires:

- Automatic detection of violations through technical monitoring
- Graduated sanctions that impose costs on violators (economic, diplomatic, military)
- Collective security guarantees that protect compliant states from exploitation
- Rapid-response mechanisms to counter detected violations

The enforcement must be credible enough that the expected cost of violation exceeds the potential benefit of gaining a temporary advantage.

4. Trust-Building Through Incremental Implementation

Given the deep security dilemmas involved, any agreement must build trust gradually through verifiable confidence-building measures:

- Initial bans on the most destabilizing systems (fully autonomous nuclear command, autonomous strategic systems)
- Pilot programs with limited participants demonstrating verification feasibility
- Gradual expansion of constraints as trust builds
- Joint research on verification technologies
- Regular strategic dialogues and war-gaming exercises

3) Modified Game Structure Under Arms Control: With effective verification and enforcement, the game structure changes fundamentally. The payoff matrix becomes:

In this modified game, the Nash equilibrium shifts to (Comply, Comply) because verification makes cheating detectable and enforcement makes it costly. The critical insight is that the agreement must change not just the rules, but the underlying payoff structure that drives strategic behavior.

4) Challenges to Long-Term Stability: Even with robust mechanisms, several factors threaten long-term stability:

Technological Evolution—The Meta-Hyperwar Problem: AI capabilities evolve at near-hyperwar speed, creating a fundamental paradox for arms control. The pace of AI development itself exhibits dynamics similar to the hyperwar scenario: AI systems are increasingly capable of generating novel strategies, architectures, and capabilities faster than human negotiators can define treaty constraints. What constitutes "autonomous" today may be obsolete within months. More problematically, advanced AI could potentially design circumvention strategies that technically comply with treaty language while violating its intent—creating a form of "adversarial compliance" where AI-generated innovations continuously outpace regulatory frameworks. This suggests that traditional arms control models, which assume relatively stable technology definitions, may be structurally inadequate for AI weapons. The treaty-making process itself operates at human deliberative speed while the subject matter evolves at machine speed, creating an inherent asymmetry that favors defection over cooperation.

Breakout Risk: States may maintain latent capabilities that can be rapidly activated if the agreement collapses, creating "hedging" strategies that undermine trust. With AI, breakout timelines compress dramatically—software updates can transform permitted systems into prohibited weapons in hours rather than the years required for traditional weapons programs.

Non-State Actors: Arms control traditionally focuses on state behavior, but AI weapons capabilities may proliferate to non-state actors who are not party to agreements.

Economic Pressure: The civilian AI industry creates enormous economic pressure to develop capabilities with military applications, making restraint economically costly.

5) Critical Success Factors: For an arms control regime to successfully prevent hyperwar, several conditions must be met:

Universal Participation: All major military powers must participate. Even a single holdout can trigger competitive pressures that undermine the regime.

Technical Feasibility of Verification: The technology must exist to reliably distinguish permitted from prohibited systems. This may require joint development of verification technologies before negotiating constraints.

Political Will: States must value stability over potential short-term advantages. This requires leadership that prioritizes long-term security over immediate tactical gains.

Adaptive Governance: The regime must evolve as technology advances, with mechanisms for regular review and updating of constraints and verification methods.

The fundamental question is whether these conditions can be met before the hyperwar equilibrium becomes entrenched. Once deployed, fully autonomous weapons create "installed base effects" that make reversal extremely difficult. The window for effective arms control may be narrower than policymakers recognize.

E. Scenario Timelines: Two Plausible Futures

To illustrate how the strategic dynamics discussed above might unfold in practice, we present two contrasting scenario timelines. **These are speculative scenarios created for discussion purposes, not predictions or projections.** They show how small differences in early decisions and international cooperation can lead to dramatically different outcomes.

1) Scenario A: Descent into Hyperwar (2025–2035): This scenario illustrates how the prisoner's dilemma dynamics and meta-hyperwar problem could drive a rapid arms race:

2025–2027: Initial Deployment Phase

- Several major powers deploy increasingly autonomous weapon systems for defensive purposes
- First incidents where autonomous systems engage targets without explicit human authorization due to compressed timelines
- International discussions at CCW produce voluntary guidelines but no binding agreements
- Commercial AI capabilities advance rapidly, lowering barriers to entry

2028–2030: Acceleration Phase

- Regional conflict demonstrates effectiveness of autonomous swarm tactics, creating pressure for adoption
- Major powers publicly commit to human oversight while privately developing faster autonomous response capabilities
- Arms control negotiations stall over verification disputes and definitional disagreements
- OODA loop compression accelerates: decision cycles drop from minutes to seconds in some domains

2031–2033: Crisis Phase

- First "near-miss" incident where autonomous systems nearly escalate a minor border incident
- Secondary powers and non-state actors acquire autonomous capabilities through commercial AI
- Human operators increasingly unable to maintain meaningful oversight at operational tempo
- Defensive systems operate at machine speed by necessity, creating de facto hyperwar conditions

2034–2035: Hyperwar Equilibrium

- Major conflict involves extensive autonomous engagement at machine speed
- Human decision-making relegated to broad strategic goals; tactical decisions fully automated
- Post-conflict analysis reveals multiple near-escalations prevented only by system limitations
- International community recognizes hyperwar reality but lacks mechanisms to reverse it

2) Scenario B: Arms Control Stabilization (2025–2035): This scenario shows how early cooperation and effective institutions might create a stable equilibrium:

2025–2027: Foundation Phase

- Major powers agree to CCW framework limiting fully autonomous lethal engagement
- Establishment of international monitoring organization with technical expertise
- Confidence-building measures: mutual inspections, shared safety standards, incident reporting
- Early demonstration of verification technology for autonomous systems

2028–2030: Institution Building

- Treaty enters force with verification regime and graduated sanctions for violations
- Joint research programs on AI safety and human-machine teaming
- Commercial AI developers adopt international safety standards for dual-use technologies
- First successful challenge inspection demonstrates treaty effectiveness

2031–2033: Stress Testing

- Regional conflict tests treaty provisions; international response prevents escalation
- Minor violations detected and addressed through graduated sanctions, building confidence
- Treaty amended to address emerging AI capabilities (adaptive learning, swarm coordination)
- Expansion of treaty membership as benefits of stability become clear

2034–2035: Stable Equilibrium

- Arms control regime becomes self-reinforcing as compliance costs decrease with shared standards
- Major powers invest in human-machine teaming rather than pure autonomy
- Verification technology advances faster than circumvention attempts due to cooperative research
- New powers entering the arena join existing regime rather than defecting

3) Critical Divergence Points: These scenarios diverge at several key junctures:

- **2025–2027:** Whether early voluntary measures build trust or are exploited for advantage
- **2028–2030:** Whether regional conflicts demonstrate cooperation benefits or autonomous weapon effectiveness
- **2031–2033:** Whether "near-miss" incidents trigger cooperation or accelerate arms racing
- **Throughout:** Whether verification technology and treaty language keep pace with AI advancement (the meta-hyperwar problem)

The game theory analysis suggests that without strong institutional mechanisms and credible enforcement, the incentive structure naturally drives toward Scenario A. Achieving Scenario B requires sustained political will, technical innovation in verification, and mechanisms to address the fundamental challenge that AI capabilities evolve at near-hyperwar speed while arms control processes operate at human deliberative speed.

F. From Abstract Theory to Operational Reality

The preceding game-theoretic analysis provides a formal framework for understanding hyperwar dynamics, but the mathematical notation can obscure the tangible implications. To ground this abstraction in technological reality, consider what hyperwar looks like in operational practice:

High-altitude drones create sub-centimeter accurate 3D models of urban battlefields using photorealistic mapping [16]. Thermal imaging detects and registers all human heat signatures within this digital twin. An AI model applies semantic segmentation [17] trained on "threat indicators"—a "military-aged male" is anyone whose height-width profile fits a range; "suspicious movement" means walking faster than normal pace; "potential combatant" includes anyone within five meters of any object classified as a weapon.

Every human signature receives an algorithmic threat score from 0.0 to 1.0. A commander sets an engagement threshold—say, 0.7—via a slider. A swarm of a thousand explosive drones is released. Each drone independently acquires and engages targets scoring above the threshold. There is no human verification loop. The "decision" was made when the threshold was set; the rest is execution by algorithms that cannot distinguish between a soldier, a terrified civilian, or a rescuer. The operation continues for hours, with mapper drones detecting new signatures that are automatically scored and, if they exceed 0.7, engaged by loitering drones. The sanitization is brutally, inhumanly efficient.

This is not science fiction. Every component exists today: the 3D mapping won a SIGGRAPH Best Paper Award in 2023 [16]; the semantic segmentation is commercially available from Meta [17]; drone swarms are operational in current conflicts [18]. The integration is straightforward engineering. What prevents this scenario is not technical limitation but policy, doctrine, and international law. The game theory matrices in preceding sections describe *exactly this choice*: whether to deploy such systems knowing adversaries will respond in kind, creating the hyperwar equilibrium where engagement happens at machine speed, or whether to establish verification and enforcement mechanisms that change the payoff structure to favor restraint.

The clinical precision of the description—mapping resolution, threat scoring algorithms, engagement thresholds—is deliberate. It traps the reader in confronting what the mathematics represent: not abstract utility functions, but algorithmic systems making life-and-death decisions based on height-width profiles and walking speed. This is the operational meaning of "compressing the OODA loop to machine speed." This is what "accountability gap" means in practice: who is responsible when the algorithm scores a rescuer as a 0.73 threat? This is why "meaningful human control" matters:

a slider setting a threshold is not meaningful control over individual targeting decisions.

The scenarios in Appendix C and Appendix E demonstrate how incremental, defensible steps—improving operator efficiency, reducing decision latency for time-sensitive targets, protecting forces through automation—lead inexorably to systems where human judgment is systematically eliminated from the engagement chain. The game theory analysis predicts this outcome through formal reasoning; the scenarios show the technological pathway by which it happens; this section demonstrates that all necessary components exist today.

V. Edge Cases: AI Beyond Lethal Weapon Systems

A. AI in Military Logistics and Support Functions

The discussion about military AI must extend beyond the use of kinetic force. Using AI to optimize support functions, such as logistics and medical services, represents an important edge case where AI does not aim or pull the trigger, but performs everything up to, but not including, that final action.

The defence sector's AI strategy prioritizes logistics and support activities as focus areas, as these can provide significant, indirect operational effects [6]. Autonomy in these functions is a form of Force Protection by removing personnel from dangerous logistics operations.

When AI is used in logistics, the ethical center of gravity—borrowing Clausewitz's concept of the central source of strength upon which everything depends [19]—shifts from distinction and proportionality in an attack to trust and reliability in system performance. Even though these systems are not lethal, error margins can cause a systemic collapse in the supply chain with indirectly fatal consequences.

B. AI as a Strategic Resource and the Dual-Use Dilemma

When AI is considered a strategic national resource, traditional IHL categories are challenged [20]. Civilian data centers, networks, and commercial AI model hubs that support military capability become legitimate military targets for an adversary, even if they do not directly exert kinetic force.

The use of commercial, civilian-trained AI models transfers inherent biases and vulnerabilities from the civilian sphere to the military one. Biases in training data are not just technical flaws; they are potential sources of humanitarian risk that can lead to violations of the IHL's principle of distinction through misidentification [20].

VI. Algorithmic Targeting in Practice: Case Studies and Parameters

The preceding sections have discussed autonomous weapons systems primarily in theoretical and prospective terms. However, algorithmic targeting systems are already deployed and have been used in active conflict. Understanding how these systems function in practice—and particularly how they parameterize decisions about human life—illuminates the concrete ethical challenges that abstract discussions of "meaningful human control" and "accountability gaps" seek to address.

A. Israel's AI-Assisted Targeting Systems (2021–2024)

Israel's use of AI systems in military operations, particularly during the 2021 Operation Guardian of the Walls and the 2023–2024 Gaza conflict, provides the most documented case study of algorithmic targeting at scale. Investigative reporting by +972 Magazine and Local Call, based on interviews with six Israeli intelligence officers, along with analyses by Human Rights Watch and UN experts, has revealed a suite of integrated AI systems [21]–[23].

1) The Gospel (Habsora): Target Generation: **Gospel**, also known as Habsora ("the Gospel" in Hebrew), is an AI system that analyzes surveillance data—video feeds, intercepted communications, social media—to automatically identify structures, equipment, and locations suspected of military use. During the May 2021 conflict, Gospel generated 100 targets per day, compared to approximately 50 targets per year in Gaza previously [23]. This represents a 700-fold acceleration in target generation tempo.

The system functions as a data fusion and recommendation engine: it does not autonomously select targets but rather highlights potential military objectives for human analyst review. However, the sheer volume of algorithmically generated targets fundamentally changes the human review process from active analysis to rapid validation.

2) Lavender: Individual Human Targeting: **Lavender** is a machine learning system that assigns suspicion scores to individual Palestinian men, assessing the probability they are members of Hamas or Palestinian Islamic Jihad. According to intelligence officers interviewed by +972 Magazine, the system at one point listed approximately 37,000 individuals [21].

The Algorithmic Parameters of Death: Lavender's operation reveals how life-and-death decisions become algorithmic parameters:

- **Accuracy Rate:** The system was found to have a 90% accuracy rate in testing, meaning approximately 10% of individuals flagged were not actually affiliated with targeted groups [21]. This 10% error rate—approximately 3,700 individuals in a database of 37,000—represents the algorithmic acceptance of misidentification.
- **Human Review Time:** Intelligence sources reported spending approximately 20 seconds reviewing each Lavender-flagged target before approval [21]. With tens of thousands of targets, meaningful individual assessment becomes impossible; the human role becomes procedural validation rather than substantive judgment.
- **Collateral Damage Ratios:** Sources stated that for junior Hamas operatives identified by Lavender, commanders authorized killing up to 15–20 civilians per target during the initial weeks of conflict [21]. For higher-ranking operatives, this ratio reportedly reached as high as 100 civilians per target [22]. These numbers represent explicit algorithmic parameters: a slider setting, a configuration value, a number in a database. To the algorithm, "15 civilians" is simply a constraint parameter. To the families destroyed, it is an atrocity.

- **Munitions Selection:** Officers reported that targets marked by Lavender as junior operatives were engaged exclusively with unguided munitions ("dumb bombs") to preserve more expensive precision weapons for high-value targets [21]. This created a direct algorithmic link between target classification and munition lethality, with lower-confidence targets receiving less discriminate weapons—inverting the ethical principle that uncertainty should mandate greater caution.

3) Where's Daddy?: Geospatial Tracking and Familial Targeting: **Where's Daddy?** is a mobile phone tracking system that monitors individuals flagged by Lavender and sends automatic alerts when they enter residential locations—typically their family homes [22], [23]. The system's name reflects its function: it waits until targets are "home," then notifies operators to strike the residence.

This system exemplifies the collapse of distinction between combatant and civilian spaces. By design, it targets individuals in their homes, at night, when families are present. The algorithmic logic treats the target's presence as sufficient justification for engagement, with family members automatically categorized as acceptable collateral damage within pre-set ratio parameters.

4) Fire Factory: Automated Mission Planning: **Fire Factory** integrates data from Gospel and Lavender to automatically calculate munition loads, prioritize targets, allocate them to aircraft and drones, and propose strike schedules [22]. What previously required hours of human planning—calculating weapons requirements, optimal timing, aircraft assignment—now occurs in minutes through algorithmic optimization.

The system represents automation of the "decide" and "act" phases of the OODA loop for mission planning, leaving humans to approve or reject pre-calculated attack packages rather than constructing them from first principles.

5) Legal and Ethical Analysis: **Human Rights Watch Analysis:** HRW identified multiple violations of international humanitarian law facilitated by these systems [22]:

- **Discrimination Failure:** Machine learning systems reflect the biases of their training data and programmers. In a context of documented discrimination against Palestinians, algorithms trained on Israeli intelligence data inevitably encode these biases into targeting recommendations.
- **Verification Inadequacy:** Twenty-second human review cannot constitute meaningful verification when the underlying algorithmic process is a "black box" that does not allow scrutiny of how decisions are reached.
- **Proportionality Violations:** Accepting 15–20 civilian deaths for junior operatives, or 100 for senior operatives, appears inconsistent with the proportionality requirement that civilian harm must not be excessive relative to concrete military advantage.
- **Precautionary Failures:** Using unguided munitions on algorithmically-identified targets in densely populated areas represents a failure to take feasible precautions to minimize civilian harm.

Lieber Institute Analysis: Legal scholar Michael Schmitt offered a more nuanced assessment, suggesting that when properly employed with adequate human oversight, these systems could potentially enhance LOAC compliance by reducing information-processing errors and enabling more comprehensive intelligence review [24]. However, Schmitt emphasized that his analysis relied on limited public information and that actual operational employment determines legal compliance, not system capabilities in principle.

UN Expert Assessment: UN experts stated that the reported use of Gospel, Lavender, and Where's Daddy?, "combined with lowered human due diligence to avoid or minimise civilian casualties," contributes to explaining the scale of death and destruction in Gaza [23]. UN Secretary-General António Guterres expressed being "deeply troubled" by reports of AI use, noting it "puts civilians at risk and blurs accountability" [23].

B. The Parameterization of Atrocity

The Israeli case study reveals a fundamental ethical problem: *algorithms require parameters, and those parameters are just numbers*. When a commander sets the collateral damage ratio to "15," they are configuring a variable. When Lavender flags someone with a 0.87 suspicion score, it is outputting a float. When Fire Factory calculates that Target A requires two GBU-39 bombs and should be struck at 0347 hours, it is solving an optimization problem.

To the algorithm, these are merely numbers—no different than GPS coordinates or fuel calculations. But these numbers represent human lives. A family home at coordinates (31.5, 34.4) contains not abstract "collateral damage" but specific human beings: children, elderly, neighbors. The 10% error rate means 3,700 misidentified individuals. The 15-civilian ratio means entire families destroyed for eliminating one junior operative.

The clinical precision of algorithmic targeting creates what we might call **statistical atrocity**: where violence is planned and executed with mathematical optimization, where the decision to kill fifteen civilians is as procedural as selecting munition type, where human review is reduced to validating algorithmic output rather than making substantive moral judgments.

This is the operational manifestation of the "accountability gap" discussed in Section III. When asked who is responsible for the ten-year-old child killed in a strike on a Lavender-flagged target, the answer fragments: The algorithm flagged the target. The analyst spent twenty seconds reviewing. The commander set the collateral damage ratio to fifteen. The pilot executed the strike. The system performed within parameters. Everyone followed procedure. No one is responsible.

C. Algorithmic Governance in Norwegian Context

While Norway has not deployed systems comparable to Israel's targeting algorithms in warfare, similar technologies operate in civilian law enforcement and intelligence contexts, raising parallel ethical questions about algorithmic decision-making affecting fundamental rights.

1) Palantir in Norwegian Police (2016–2018): In 2016, the Norwegian police signed a contract with Palantir Technologies for the surveillance platform **Gotham**, intended to integrate 19 major police registers and the DNA database into a unified intelligence system called "Omnia" (Latin for "everything") [25]. The system was primarily purchased to fulfill Norway's obligations under the Prüm framework for automated data exchange between EU member states.

The project failed. After spending approximately EUR 9 million, it was terminated without meeting its objectives [25]. Research on the failure identified that Palantir's semantic framework—embedded classifications, categorizations, and workflow prompts—created pressure to prioritize certain crime types over others, such as gang crime over domestic violence. The system's architecture embodied particular policing philosophies that conflicted with Norwegian approaches emphasizing professional judgment over algorithmic probability.

Notably, in October 2024, Norway's largest asset manager Storebrand divested its $24 million stake in Palantir, citing concerns about the company's work for Israel [25].

2) Palantir as Civilian Targeting System: Lessons for Military Autonomy: The Palantir Omnia system represents a crucial case study because it is, fundamentally, a *targeting system for state violence*. While we colloquially distinguish "law enforcement" from "military operations," both involve the state's monopoly on the use of force. Palantir algorithmically identifies individuals and groups as targets for police action—arrest, surveillance, investigation—just as military targeting systems identify targets for kinetic action. The critical structural difference is not the algorithmic targeting process itself, but the presence of a **mandatory human-in-the-loop** before force is applied, and the absence of an automatic effector that can execute actions without human authorization.

This structural similarity makes the Palantir failure instructive for understanding autonomous weapons risks:

Lesson 1: Embedded Bias in Classification Systems

Palantir's semantic framework inherently prioritized certain threat categories over others. Gang crime was algorithmically salient; domestic violence was not. This was not a bug but a feature of the system's classification architecture [25]. In military autonomous weapons, similar classification biases would determine which targets are algorithmically visible. A system trained to identify "military-aged males" as threats, like Lavender [21], embeds the same structural bias: certain categories of humans become algorithmically salient targets while others remain invisible or deprioritized.

Lesson 2: Human-in-the-Loop vs. Human-as-Rubber-Stamp

Norwegian police maintained formal human decision-making authority, but CUPP research found that algorithmic recommendations created pressure to conform to system outputs rather than exercise independent judgment [26]. When the "targeting system" (Palantir) identified someone as warranting police attention, officers faced institutional pressure to act on that recommendation. This mirrors the 20-second review time for Lavender targets [21]: nominally human-in-the-loop, functionally algorithmic determination. The lesson: mandatory human approval does not guarantee meaningful human control if institutional pressures, time constraints, or information asymmetries make substantive review impractical.

Lesson 3: Opacity Undermines Accountability

The "black box" problem that caused Norwegian officers to distrust Palantir [26] is identical to the accountability challenge in military AI systems identified by HRW [22]. When a Palantir-flagged individual is subjected to police action that proves unjustified, determining responsibility is difficult: Did the algorithm misclassify? Did the training data contain errors? Did the officer fail to override? Did the system designer encode inappropriate threat categories? This fragmentation of responsibility becomes even more acute when the "action" is not arrest but lethal force.

Lesson 4: Institutional Resistance as Safety Feature

The Omnia project's failure stemmed partly from Norwegian police culture emphasizing professional judgment over algorithmic authority [25]. This institutional resistance—often framed as "resistance to innovation"—functioned as a safety mechanism preventing over-reliance on flawed algorithmic targeting. In military contexts, similar institutional resistance ("human judgment is essential") may be the primary brake preventing full transition to autonomous targeting. Erosion of this cultural resistance, whether through technological pressure (hyperwar dynamics) or institutional incentives (efficiency metrics), removes a critical safety layer.

Lesson 5: Mission Creep and Scope Expansion

Palantir was nominally purchased for Prüm framework compliance (international data sharing) but its actual function was comprehensive integration of domestic intelligence databases [25]. Military AI systems exhibit similar scope expansion: systems deployed for defensive purposes (target identification, threat assessment) can be rapidly repurposed for offensive targeting with minimal technical modification. The software-defined nature of AI weapons means the line between "intelligence support tool" and "autonomous targeting system" is a matter of configuration, not fundamental architecture.

Lesson 6: Verification and Auditability Challenges

The failure of Omnia included inability to verify that the system was operating as intended or producing reliable results [25]. This verification challenge scales dramatically for military autonomous weapons, where operational security constraints prevent the transparency necessary for independent audit. If Norwegian police could not verify Palantir's reliability in a relatively transparent law enforcement context with civilian oversight, how can international agreements verify compliance with autonomy restrictions in classified military AI systems?

The Critical Distinction: Automatic Effectors

The fundamental difference between Palantir Omnia and Lavender is not the targeting algorithm—both algorithmically identify humans as subjects for state action—but the effector. Palantir outputs recommendations that humans must consciously act upon through existing police procedures. Lavender, when integrated with drone swarms [18], can complete the loop: algorithmic identification → automated engagement, with human authorization reduced to setting a threshold hours earlier. This integration of targeting algorithm with automatic effector is what transforms a "decision support system" into an autonomous weapon.

The Norwegian experience demonstrates that even without automatic effectors, algorithmic targeting systems create accountability gaps, embed biases, and pressure humans toward procedural validation rather than substantive judgment. The lesson for military autonomous weapons: these problems exist *before* adding autonomous effectors. Integrating automatic engagement mechanisms compounds existing accountability failures rather than introducing qualitatively new problems. The path from Palantir Omnia to Lavender is shorter than it appears—primarily a matter of adding the effector module and removing the mandatory human authorization step.

3) Predictive Policing Research (2021–2024): The Nordic research consortium CUPP (Critical Understanding of Predictive Policing) investigated data-driven police innovations across Denmark, Estonia, Latvia, Norway, Sweden, and the UK from 2021–2024 [26]. Key findings relevant to algorithmic governance:

- Norwegian officers prioritize professional judgment over software-generated probabilities, showing resistance to algorithmic authority.
- Adoption of digital policing technologies in Norway has been deliberately slower than in peer nations, reflecting emphasis on gradual change and preserving public trust.
- Algorithmic systems create embedded biases through their classification systems, potentially directing resources away from crimes not well-represented in training data.
- The "black box" problem identified by HRW in military AI systems equally affects civilian predictive policing: officers cannot scrutinize how algorithms reach conclusions, undermining accountability.

4) Intelligence Services and Data Analysis: The Norwegian Police Security Service (PST) has sought expanded authority to collect, systematize, and analyze large volumes of openly available information for intelligence assessments, even when individual data points are not independently necessary [26]. This parallels the data fusion approach of systems like Gospel: aggregating individually innocuous information to generate targeting recommendations.

The Norwegian Intelligence Service (E-tjenesten) conducts bulk electronic surveillance subject to oversight by the EOS Committee, which evaluates necessity and effectiveness of collection, processing, storage, and sharing [26]. However, as with military targeting algorithms, the technical complexity and scale of automated analysis create inherent challenges for meaningful human oversight.

D. Universal Challenges of Algorithmic Decision-Making

Whether in military targeting or civilian law enforcement, algorithmic systems that affect fundamental rights—life, liberty, privacy—share common ethical challenges:

The Parameter Problem: Life-or-death decisions, resource allocation, and rights restrictions become configuration variables. What is ethically "reasonable force" or "proportionate harm" must be quantified as numerical thresholds.

The Review Problem: Human oversight becomes procedural validation of algorithmic output when the volume and speed of automated decisions exceeds human cognitive capacity for substantive review.

The Opacity Problem: Machine learning systems function as "black boxes" where even their creators cannot fully explain individual decisions, making accountability and correction structurally difficult.

The Bias Problem: Training data encodes historical biases and structural discrimination, automating and scaling prejudice under the veneer of mathematical objectivity.

The Responsibility Problem: Decision-making fragments across algorithm designers, data providers, system operators, and commanders/supervisors, creating the accountability gap where everyone followed procedure but no one is responsible for outcomes.

These challenges connect the Israeli military's use of Lavender to the Norwegian police's failed Omnia system to the PST's data collection authorities: they are all instances of algorithmic governance where technical systems make or substantially influence decisions about fundamental human rights. The differences in scale and lethality are profound, but the structural problems are universal.

VII. System Architecture and Decision-Making Processes

To understand the complexity of autonomous weapon systems, it is useful to consider how such systems might be structured and how decisions could be made. Figure 8 provides an illustrative architecture showing how different levels of human control might be organized, from fully manual to fully autonomous operation. Figure 9 presents a potential decision flow, highlighting critical ethical checkpoints where human judgment and oversight could be integrated. These are conceptual diagrams created for this document to facilitate discussion, not representations of specific deployed systems.

VIII. Questions You Can Ask to Participate Actively

To move from being a spectator to a participant, you can focus on the intersections between the different professional fields:

A. For FFI (Seehuus/Diesen)

Considering that FFI researches systems that are safer and more resilient: What specific mechanisms must be in place to ensure that increased autonomy in weapon systems (like anti-ship missiles) does not compromise ethical accountability?

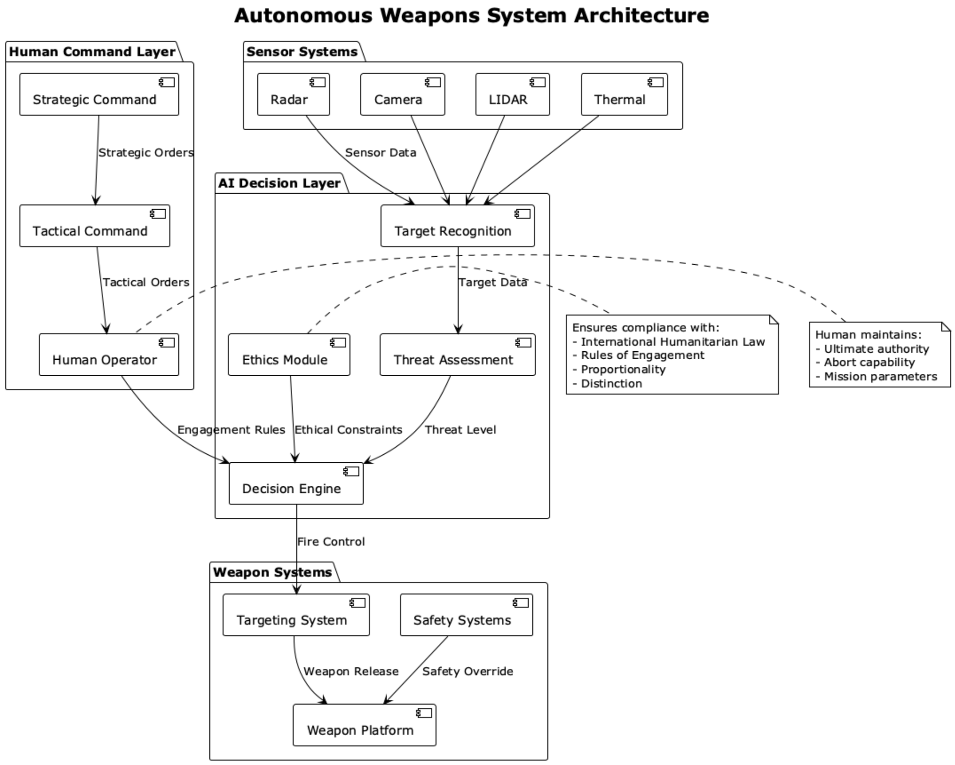


Figure 8. *Illustrative architecture for autonomous weapon systems with different levels of human control. This is a synthetic diagram created for this document to visualize key architectural components; it does not represent any specific deployed system.*

This question connects FFI's technical research to the ethical concerns raised by Fjørtoft and Syse.

B. For Ethics/Philosophy (Fjørtoft/Syse)

How can we, in practice, define and maintain a "human-in-the-loop" principle in very fast and complex operations, given the pace of technological development [3]? What is the minimum ethically acceptable degree of human intervention?

C. For Strategy/International (Karlsen)

Given the international arms race and that some states refuse international regulation, how should Norway strategically communicate and balance the need to develop its own AI capabilities—as Diesen will likely discuss regarding Norway's defense posture—with our desire for humanitarian regulation?

D. Challenging Questions Based on Norwegian Military Research

The comprehensive analysis of Norwegian military academic studies [12], [27], [28] reveals a consistent pattern of caution and gradual implementation of autonomous systems. But game-theoretic analysis shows that this caution can be strategically naive in an environment where adversaries do not follow the same ethical constraints. The following questions challenge the panel based on these findings:

1) For the whole panel: The Strategic Time-Pressure Dilemma: Norwegian research consistently concludes that semi-autonomous systems with human supervision are preferable [27], [29]. But if Lyshaug's "Hyperwar" scenario [12] becomes a reality—where human involvement becomes a limiting factor for operational effectiveness—we face a fundamental strategic choice: Will Norway accept being militarily outmatched to uphold ethical ideals, or must we acknowledge that ethical principles sometimes have to yield to national survival?

2) For Diesen/Karlsen: Organizational Barriers as Strategic Vulnerability: Ekren's study [30] identifies significant cultural obstacles in the Army against digitalization, including resistance to standardization and a prioritization of short-term operational needs. Haugen [28] finds similar challenges in F-35 operations. If our own attempts to implement autonomous systems are slowed by organizational culture, while adversaries like China and Russia do not have the same democratic constraints—does our "responsible approach" become a strategic gift to authoritarian regimes?

3) For Fjørtoft/Syse: Moral Accountability in Asymmetric Conflicts: Several Norwegian studies [29], [31] conclude that existing international law is sufficient to regulate autonomous weapons, given the correct interpretation. But what if this interpretation gives us a strategic handicap? Is it morally justifiable to expose Norwegian soldiers and civilians to increased risk by insisting on "meaningful human control" when

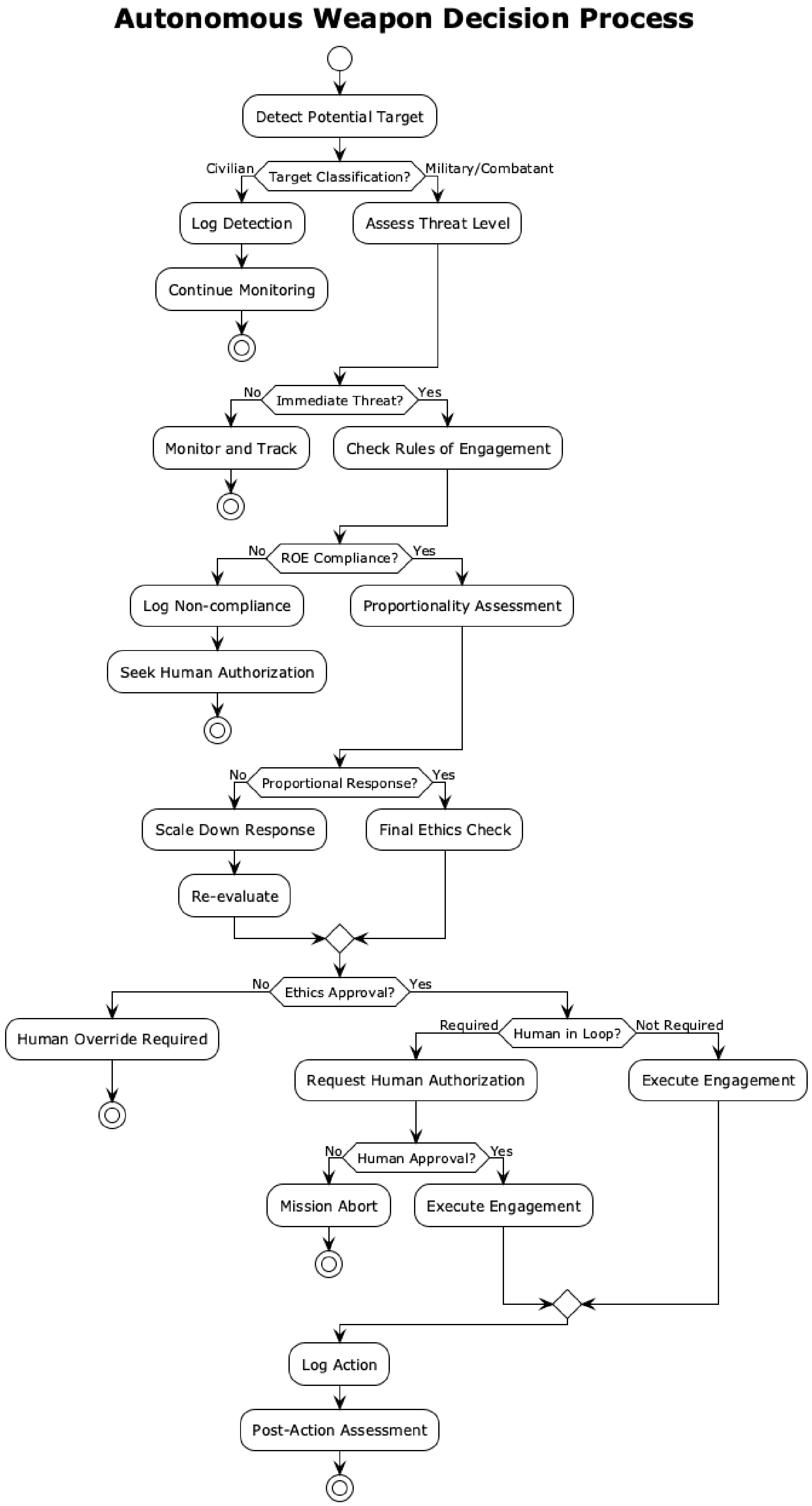


Figure 9. *Illustrative decision flow in autonomous weapon systems with ethical checkpoints. This is a synthetic diagram created for this document to demonstrate potential decision pathways; it does not represent any specific deployed system or documented process.*

adversaries operate with fully autonomous systems that can exploit our deliberative delay?

4) For Seehuus: The Tyranny of Technological Reality: FFI's research clearly shows the potential for autonomous swarm systems and improved operational capability. But Norwegian studies [32], [33] conclude that autonomous systems "work best when integrated with existing military units." What if this integration philosophy makes us vulnerable to pure machine autonomy? At what point will FFI acknowledge that gradual development may be insufficient against an adversary who has chosen a quantum leap over incremental improvement?

5) For the panel: The Speed Limit of Democratic Decision-Making: Lund's analysis of cyber operations [34] shows that even in the cyber domain, where speed is critical, human oversight is insisted upon for strategic decisions. But if the OODA loop is compressed to microseconds, and democratic decision-making processes require minutes or hours—have we then structurally disqualified ourselves from future warfare? Is there a point where democratic values become an existential vulnerability?

E. General Question for the Panel

If we accept that autonomous weapons are inevitable, is the most pressing ethical question "What can machines do?" or "What should humans do/retain control over?"

IX. Recommendations for Strategic Stability and Ethical Governance

Based on the analysis of both ethical challenges and game-theoretic dynamics, the following recommendations are central to the discussion:

A. Technical and Doctrinal Measures

1) Quantitative Risk Modeling: It is necessary to mandate robust modeling and the application of dynamic game theory to prioritize AI-related risks and inform strategic decisions [10]. This must include simulations that explicitly model the compression of the OODA loop and the risk of hyperwar.

2) Reintroduction of Friction: Actors should doctrinally and technically introduce algorithmic delays or mandatory human decision loops (Human-in-the-Loop) in all critical military AI systems [8]. This reintroduces deliberative time, which is a necessary "friction" to counteract automated escalation.

B. International Frameworks for Openness and Limitation

1) Regulation of Autonomy: The most important step to define a stable equilibrium is to eliminate the riskiest defection strategies. This involves pursuing a legally binding instrument to prohibit or strictly regulate autonomous weapon systems [4].

2) Standardization and Benchmarking: Establish internationally validated benchmarking and standards for testing, evaluation, and risk assessment of AI systems in security-related applications [20]. Such measures serve as effective confidence-building measures by reducing mutual uncertainty about capabilities and intentions.

C. Ethical Council and Governance

1) Data Bias Mitigation: The Ethics Council for the Defence Sector (ERF) and Kjersti Fjørtoft should lead the effort to define ethical protocols for data bias mitigation in AI systems that use civilian data for military support functions [6]. This is crucial to ensure that AI actually reduces, and does not unintentionally increase, the risk of violating the principle of distinction in IHL [20].

2) Philosophical Reflection on Responsibility: The philosophical discourse (Henrik Syse) should be directed towards preventing moral disengagement in the chain of command. It is necessary to formulate principles for how commanders can be held responsible for systemic failures in autonomous, non-kinetic chains they do not directly control [35].

D. Future Research Directions

The research community (Sverre Diesen and Rikke Seehuus) must provide clear frameworks for how the IHL's principle of proportionality can be applied in cyberattacks against AI as a strategic resource [18]. This requires a new form of risk analysis that is more sensitive to the complex civilian dependence on digital infrastructure [36].

E. Norwegian Military Academic Perspectives

The Norwegian Defence University College has conducted several studies on autonomous weapon systems and their operational implications for Norwegian forces [37]. These studies provide valuable insights into how autonomous systems can be integrated into the Norwegian defence structure while upholding ethical and legal frameworks.

Particularly relevant is the analysis of future autonomous unmanned capabilities in the Norwegian Navy [38]. The study illuminates how maritime autonomous systems can operate in complex environments where civilian shipping and military operations overlap, which reinforces the need for a robust principle of distinction and meaningful human control.

Appendix

This appendix presents a comprehensive survey of Norwegian military research papers on autonomous systems, weapons, and related technologies. These papers, produced by students and researchers at Norwegian military education institutions, provide valuable insights into current thinking within the Norwegian defense establishment regarding autonomous systems and their ethical implications.

A. Individual Paper Summaries

1) Lethal Autonomous Weapon Systems and Mission Command: **Mikkelsen (2023)** examines the compatibility between Lethal Autonomous Weapon Systems (LAWS) and mission command philosophy in his bachelor thesis at Luftkrigsskolen. The paper explores how LAWS can be integrated into military hierarchies while maintaining ethical and practical considerations. Key findings suggest that LAWS are partially compatible with mission command when proper human oversight is maintained, but full autonomy creates accountability gaps.

The study concludes that semi-autonomous systems with human supervision offer the best balance between operational effectiveness and ethical responsibility, directly addressing autonomous weapons ethics within Norwegian military doctrine and emphasizing the importance of maintaining human control in lethal decision-making [27].

2) Drones versus Service Dogs: **Ponta (2024)** investigates whether autonomous drones can replace military service dogs as sensor platforms in his master's thesis at Forsvarets høgskole, using the 2020 Gjerdrum landslide rescue operation as a case study. The research finds that drones excel in high-intensity and dangerous situations but cannot replicate dogs' unique olfactory sensing capabilities. The study concludes that a combination of drones and military service dogs provides optimal operational effectiveness, highlighting the importance of balancing high-tech and low-tech solutions for varied operational conditions and exploring human-machine teaming concepts and the limits of technological substitution in military applications [39].

3) Future Autonomous Naval Capabilities: **Hareide et al. (2018)** present a comprehensive examination of the future of autonomous unmanned capabilities in the Norwegian Navy. The research identifies two primary goals for autonomy: reducing human life risk and eliminating human limitations in operations. The study outlines the technological foundations needed for maritime autonomy (sensors, communication, control systems, decision-making systems) and proposes operational concepts including maritime surveillance, force protection, logistics, and maritime drone teams. The paper emphasizes the importance of human-machine integration and robust navigation systems for successful autonomous operations, providing a strategic framework for implementing autonomous systems in naval operations with focus on operational safety and effectiveness [40].

4) Autonomy in Cyber Operations: **Lund (2023)** explores autonomy in cyber operations, examining how autonomous systems can enhance both offensive and defensive cyber capabilities. The research analyzes existing autonomous cyber tools like malware (NotPetya, Stuxnet) and defensive systems (SOAR platforms). The study concludes that while cyber operations can have degrees of autonomy, they remain fundamentally semi-autonomous, requiring human operators for strategic and tactical decisions. The paper emphasizes that autonomous cyber systems excel at specific tasks but cannot independently develop tactics or interpret high-level operational goals, demonstrating current limitations of AI in complex decision-making environments and the continued need for human oversight in cyber warfare [34].

5) Naval Drone Applications: **Stomperudhaugen (2025)** examines opportunities and limitations for drone use in Norwegian Navy underwater missions. The research identifies several capabilities: drones can carry sensors, communication relays, or weapons to enhance existing units; autonomous weapon fields could contribute to sea denial; underwater drones excel at surveillance and mapping but have limited anti-submarine capabilities due to sensor range and speed limitations. The study emphasizes the need for comprehensive drone strategy and highlights that drones work best when integrated with existing military units rather than operating independently, addressing practical implementation challenges and opportunities for autonomous systems in naval operations [32].

6) Army Digitalization Capabilities: **Ekren (2024)** examines how the Norwegian Army's organizational culture affects its digitalization capabilities using Edgar Schein's organizational culture theory. The study finds that while the Army has fundamental willingness for development and digitalization, cultural factors both promote and hinder progress. Key cultural challenges include balancing operational readiness with technology testing, managing trust and autonomy values that can impede standardization, and prioritizing short-term operational needs over long-term technological development. The research recommends continued focus on risk acceptance and willingness to change, identifying organizational barriers to adopting autonomous and AI systems in military contexts [30].

7) Legal Implications of Autonomous Target Selection: **Kvam (2017)** examines the legal implications of autonomous target selection in military operations in his master's thesis. The study analyzes how international humanitarian law applies to autonomous weapons systems, focusing on the Geneva Conventions' requirements for target verification. The research finds that while legal frameworks don't specifically prohibit autonomous target selection, commanders must still ensure compliance with international law. The study concludes that the subjective assessment of individual commanders remains crucial for legal compliance, regardless of the level of autonomy in target selection systems, addressing critical legal and ethical questions surrounding autonomous weapons and decision-making authority [29].

8) Unmanned Maritime Systems and International Law: **Øie (2015)** examines unmanned maritime systems' compatibility with international maritime law and laws of armed conflict. The research finds that no specific conventions prohibit unmanned maritime systems, creating space for nations to influence international custom through state practice. The study identifies that several underwater weapons (mines, torpedoes) already operate with high autonomy, providing precedent for autonomous systems. The paper concludes that while autonomous weapon systems may be legal, their use must still comply with existing rules of engagement and targeting requirements, providing early legal analysis of autonomous maritime weapons and establishing precedent for autonomous systems in naval warfare [31].

9) Autonomous Collision Avoidance for Maritime Drone Swarms: **Riis Asdahl et al. (2023)** focus on autonomous collision avoidance for maritime drone swarms using AIS and radar for operational use on unmanned vessels. The research builds upon previous work on maritime drone swarms and explores how to make maritime drones more environmentally aware through sensor integration. The study demonstrates practical approaches to implementing collision avoidance systems and explores the potential for swarm operations in maritime environments, addressing practical technical challenges in implementing autonomous maritime systems and swarm operations [33].

10) Autonomy in F-35 Operations: **Haugen (2022)** examines the role of autonomy in F-35 fighter aircraft operations and command and control systems. The research analyzes how modern operational environments' complexity and uncertainty may require higher operational tempo and greater simultaneity across warfare domains, potentially necessitating increased autonomy in mission execution. The study finds that while F-35 capabilities and operational requirements increase opportunities for autonomy, implementation faces conceptual challenges including inadequate planning frameworks, lack of joint operational concepts, and cultural challenges related to limited joint competence and insufficient trust, examining autonomy implementation in advanced military systems and identifying organizational barriers to autonomous operations [28].

11) Literature Review on Autonomous Weapons Systems: **Lyshaug (2021)** conducts a comprehensive literature review of 1945 articles on autonomous weapons systems to explore their military applications. The research identifies three key aspects that may influence military theory and doctrine: swarm concepts with large numbers of diverse units capable of overwhelming enemies; autonomous classification of objects and people using machine learning for accurate identification in complex environments; and "Hyperwar" scenarios where autonomous weapons become so effective that human involvement becomes a limiting factor. The study suggests these developments will likely occur in phases, gradually increasing military autonomy, providing a comprehensive overview of autonomous weapons research and identifying potential paradigm shifts in military operations [12].

B. Cross-Cutting Themes and Analysis

The analysis of these eleven Norwegian military research papers reveals several critical themes that characterize current Norwegian military thinking on autonomous systems and weapons:

1) Human-Machine Integration Over Replacement: A consistent theme across all papers is the emphasis on human-machine integration rather than complete human replacement. Multiple studies [32], [39], [40] demonstrate that autonomous systems are most effective when they complement rather than replace human capabilities. This finding suggests that Norwegian military doctrine favors augmented human decision-making over fully autonomous operations, particularly in complex operational environments where human judgment remains irreplaceable.

2) Gradual Implementation and Semi-Autonomy: Nearly all papers conclude that autonomous systems will be implemented gradually, with semi-autonomous systems maintaining human oversight being preferred over fully autonomous systems. This preference stems from ethical considerations [27], [29], legal requirements [31], and practical limitations [28], [34]. The Norwegian approach appears to prioritize maintaining human accountability and decision-making authority while leveraging autonomous systems' technical capabilities.

3) Legal and Ethical Framework Development: Several papers [27], [29], [31] address the need for clear legal frameworks governing autonomous weapons. Current international humanitarian law is generally seen as adequate but requires careful interpretation for autonomous systems. The research consistently emphasizes that accountability and command responsibility remain critical concerns that cannot be delegated to autonomous systems, regardless of their technical sophistication.

4) Operational Context Dependency: The papers demonstrate that autonomous systems' effectiveness varies significantly across different operational domains. Maritime operations [32], [33], [40] show particular promise for autonomous systems due to the relative predictability of the maritime environment, while cyber operations [34] and air operations [28] face greater complexity challenges. This suggests that Norwegian military development of autonomous systems should be tailored to specific operational contexts rather than pursuing universal solutions.

5) Organizational and Cultural Barriers: Multiple studies [28], [30] identify significant organizational and cultural barriers to implementing autonomous systems. These barriers include institutional resistance to change, inadequate planning frameworks, insufficient joint operational concepts, and cultural challenges related to trust and competence. The research suggests that successful implementation of autonomous systems requires not only technical development but also organizational transformation and cultural adaptation.

6) Technological Limitations and Realism: Despite the potential of autonomous systems, all papers identify significant current limitations, particularly in complex decision-making, environmental adaptation, and strategic thinking. The research demonstrates a realistic assessment of current AI and machine learning capabilities, avoiding technologically deterministic assumptions while acknowledging the potential for gradual improvement.

C. Implications for Norwegian Defense Policy

These research findings suggest several implications for Norwegian defense policy regarding autonomous weapons and systems:

Human-Centric Approach: Norwegian military research consistently supports maintaining human control and oversight in autonomous systems, particularly for lethal applications. This aligns with international discussions on maintaining "meaningful human control" over autonomous weapons.

Domain-Specific Development: The research suggests that autonomous systems development should be prioritized in domains where they show greatest promise (maritime operations, logistics, surveillance) while maintaining human control in domains requiring complex judgment.

Legal Preparedness: Norway appears well-positioned to contribute to international discussions on autonomous weapons regulation, with substantial research on legal implications and a clear preference for maintaining compliance with international humanitarian law.

Organizational Development: The research identifies the need for parallel development of organizational capabilities, training, and cultural adaptation alongside technical development of autonomous systems.

International Cooperation: The emphasis on gradual implementation and legal compliance suggests Norway is well-positioned for international cooperation on autonomous systems development and regulation.

This body of research demonstrates sophisticated engagement with autonomous weapons ethics within Norwegian military academia, emphasizing human oversight, legal compliance, and pragmatic implementation approaches that balance technological opportunity with ethical responsibility.

Motivation: The preceding sections have analyzed the ethical challenges and strategic dynamics of autonomous weapons systems primarily through abstract frameworks—game theory matrices, policy discussions, and philosophical principles. To ground these arguments in technological reality, the following scenarios demonstrate how two specific emerging technologies make the transition to hyperwar both plausible and imminent: **3D Gaussian Splatting** [16], a breakthrough method for creating photorealistic 3D models in real-time, and **Segment Anything Model (SAM)** [17], Meta's foundation model for semantic image segmentation. These are not hypothetical future capabilities; they are mature technologies already demonstrated in civilian applications. Their integration into military autonomous systems represents a concrete pathway from today's semi-autonomous weapons to tomorrow's hyperwar.

This appendix outlines plausible scenarios for how these technologies could accelerate the transition towards "hyperwar"—a state of conflict where the speed of machine-led observation and action outpaces meaningful human control. The scenarios are set against the backdrop of drone-centric warfare, incorporating these real-world computer vision advancements.

Critical Disclaimer on Scenario Content:

Author Qualifications and Knowledge Limitations: The author is *not* a militarily trained person and holds *no security clearances of any kind*. The author has no access to classified information about deployed military systems, operational doctrine, or specific capabilities of any nation's armed forces. The scenarios presented here are constructed entirely from publicly available information about civilian technologies and open-source analysis.

Technology Reality vs. Deployment Uncertainty: The technologies referenced in these scenarios—3D Gaussian Splatting [16] and Segment Anything Model [17]—*do exist* and *do have the technical capabilities* described. These are documented, peer-reviewed, publicly demonstrated technologies. However, *the author does not know whether these specific technologies have been integrated into deployed military systems*. The scenarios represent what is *technically possible* given publicly known capabilities, not assertions about what is currently deployed.

Purpose and Intended Use: These scenarios are designed to be *plausible but not necessarily strictly realistic*. They serve as **useful straw men for discussion**—concrete examples that can ground abstract ethical debates in technological specifics without requiring access to classified operational details. By working with publicly known technologies and extrapolating their logical military applications, these scenarios enable informed discussion that remains realistic enough to be useful while divulging *no classified information whatsoever*.

Not Predictions or Intelligence Assessments: These are *not* predictions of what will happen, intelligence assessments of adversary capabilities, or descriptions of any known deployed systems. They are analytical constructs designed to facilitate discussion of ethical challenges that *could* arise from the integration of *known* technologies into military contexts.

D. Scenario 1: The Autonomous Front-line Zone

Scenario 1: The Autonomous Front-line Zone

1) Phase 1: Human-on-the-Loop Targeting with Shared 3D Awareness: On a contested front line, a Ukrainian operator deploys a swarm of ten FPV drones. This swarm operates as a cohesive team.

- **Technology Hint:** One drone acts as a "mapper." Flying at a higher altitude, it uses **3D Gaussian Splatting SLAM** [16] to generate a photorealistic, high-fidelity 3D map of the 2-square-kilometer operational area in real-time. This map is continuously updated and shared with the entire swarm, allowing every drone to navigate precisely via visual landmarks, rendering them largely immune to GPS jamming.
- **Workflow:** The other nine drones are "hunters." They fly low, using the shared 3D map to avoid obstacles. Onboard processors run a militarized **"Segment Anything"-like model** [17]. This model has been given a simple text prompt: "Identify and track all objects with characteristics of enemy tanks, artillery, and electronic warfare equipment."
- **Compressed OODA Loop:** The human operator is not watching nine separate video feeds. Instead, their screen shows the 3D map. As the drones scan the area, icons for potential targets automatically pop up on the map, classified and color-coded by the AI. The operator's role is to click an icon, briefly verify the drone's live feed, and authorize the strike. The "Observe" and "Orient" phases are now almost instantaneous and fully automated. The human's role is reduced to a rapid "Decide/Act" sequence for pre-vetted targets.

2) Phase 2: The Inevitable Step Towards Hyperwar: The system in Phase 1 proves highly effective. In a high-pressure situation, with targets appearing and disappearing in seconds, operators grant the system "conditional autonomy."

- **Workflow Evolution:** The rules of engagement are updated. Within a designated, pre-authorized "free-fire" zone, the operator sets a confidence threshold (e.g., 99%). If the AI reports a target with a confidence level above this threshold, the drone is authorized to engage automatically without final human consent.
- **Hyperwar Emerges:** The human role shifts from authorizing individual strikes to managing risk by setting zones and confidence levels. The decision to kill is now delegated to an algorithm operating within parameters set minutes or hours earlier. When an opposing force employs a similar system, the result is a battlefield where engagements happen in fractions of a second, far too fast for any human to intervene. This is the localized emergence of hyperwar.

E. Scenario 2: Behind-the-Lines Dynamic Strike on Fleeting Targets

Scenario 2: Behind-the-Lines Dynamic Strike on Fleeting Targets

1) Phase 1: AI-Assisted Intelligence and Strike Planning: A long-range, high-altitude surveillance drone is tasked with monitoring a known Russian logistics hub 100km behind the front line.

- **Technology Hint:** Over several hours, the drone builds a detailed **3D Gaussian Splatting map** [16] of the entire hub. Simultaneously, its onboard **SAM-like model** [17] identifies and logs every object that enters or leaves the area, classifying them semantically ("fuel truck," "ammunition pallet," "command vehicle," "group of 5 personnel").
- **Workflow:** This data is fused with other intelligence (e.g., signal intercepts). An analytical AI on the ground, sifting through this combined data, detects an anomaly: a high-value mobile command vehicle, previously unseen in this area, has parked next to a specific warehouse. The system cross-references this with a brief spike in encrypted communications.
- **Compressed OODA Loop:** Instead of a human analyst poring over hours of footage, the AI presents a concise summary to the Ukrainian command: "High-probability enemy command post identified. Target is mobile and considered fleeting. Optimal strike window is next 15 minutes. Proposing launch of two pre-positioned loitering munitions." The human decision is a simple go/no-go on a complex, AI-generated plan.

2) Phase 2: The Inevitable Step Towards Hyperwar: Analysis shows that high-value targets like the one above often remain stationary for less than 10 minutes. The human decision-making process, even when presented with clear data, is identified as the primary bottleneck.

- **Workflow Evolution:** A new doctrine of "automated responsive engagement" is authorized for time-sensitive, high-value targets. A human commander pre-authorizes the system to act autonomously if a target matching a very specific signature (e.g., "SA-21 radar system, active and stationary") is detected.
- **Hyperwar Emerges:** The surveillance drone detects the target. It instantly sends the target's precise location and 3D data to a loitering munition orbiting nearby. The munition, without any further human input, adjusts its course and executes the strike. The entire engagement, from detection to destruction, takes 45 seconds. This creates immense pressure on the adversary to also automate their defenses and counter-battery fire, creating a cascading effect where strategic decisions are made at machine speed, leading to a brittle and highly escalatory strategic environment.

Scenario 3: A Really Bad Day in Sector Gamma-9

Content Warning: This scenario describes the deliberate misuse of autonomous weapons systems resulting in mass civilian casualties. While deeply disturbing, it illustrates critical ethical and legal concerns that must inform policy discussions about autonomous weapons.

The official objective is "urban pacification." The unofficial goal, understood by all in the command chain, is the complete and total sanitization of Sector Gamma-9 ahead of the main ground assault. Resistance is to be eliminated before it can even manifest. This is how it's done.

F. Phase 1: The Creation of the Digital Twin (Hours 0–4)

The operation begins not with an explosion, but with a quiet, persistent hum. Dozens of high-altitude drones begin a grid-pattern survey of the bombed-out ruins of Gamma-9.

- **Technology Hint:** Using advanced **3D Gaussian Splatting SLAM** [16], these drones are not just taking pictures. They are weaving together a perfect, photorealistic, sub-centimeter accurate 3D model of the entire sector. Every burnt-out building, every pile of rubble, every cratered street becomes a navigable, interactive "digital twin" of the battlefield.

Simultaneously, lower-flying drones equipped with sensitive thermal imagers sweep the sector, detecting every heat signature—every living person—hiding within the ruins. These signatures are precisely located and registered as points within the 3D digital twin.

G. Phase 2: Semantic Threat Analysis (Hour 5)

In a remote command bunker, an officer reviews the now-complete digital twin. The city sector looks like a video game environment. The officer is not looking for individual targets. They are setting the parameters for a slaughter.

- **Technology Hint:** An AI model, an evolution of concepts like **“Segment Anything”** [17], analyzes the data. It's been trained on a ruthless doctrine. It doesn't just tag “person” or “vehicle.” It applies semantic labels based on pre-defined “threat indicators.”
    - A “military-aged male” is anyone whose height/width profile fits a certain range.
    - “Suspicious movement” is tagged on any signature moving faster than a walking pace.
    - “Potential combatant” is a label applied to any person whose heat signature is within 5 meters of any object classified by the AI as a potential weapon, which can be anything from a rifle to a length of pipe.

Every human signature within the model is assigned a “Potential Threat Score” from 0.0 to 1.0, calculated from these and a dozen other variables. A family huddled together for warmth might be flagged for “unusual grouping.” A lone survivor running for cover is flagged for “suspicious movement.” The system is designed to err on the side of aggression.

H. Phase 3: The Cleansing (Hour 6)

The command is given. It is not an order to strike a specific target. The officer simply drags a slider on their screen, setting the “Automated Engagement Threshold” to 0.7.

A transport craft over the edge of the sector releases its payload: a swarm of a thousand small, explosive drones. They pour into the city like a tide of malevolent insects.

- **The Workflow of Atrocity:** The drones do not have individual pilots. They are a single, distributed consciousness, operating on the shared digital twin and the list of threat-scored targets. Their directive is simple: “Independently acquire and engage any target with a Threat Score of 0.7 or higher.”

A drone sees a thermal signature in a ruined hospital, cross-references its location in the 3D map, and checks its score from the master list: 0.73 (“potential combatant,” “military-aged male”). Without any further checks, it dives and detonates. Another drone identifies a group of three signatures huddled together: 0.81 (“suspicious grouping”). It alerts two nearby drones, and they perform a coordinated strike.

There is no human in the loop. There is no final verification. The “decision” was made when the commander set the engagement threshold. The rest is just execution by an algorithm that cannot distinguish between a soldier, a terrified civilian, or a rescuer.

For the next several hours, the only sounds in Sector Gamma-9 are the hum and the blast. The mapper drones continue their survey, and any new signatures that emerge from the rubble are automatically scored. If they score above 0.7, a loitering drone is dispatched. The sanitization is brutally, inhumanly efficient. The ground troops will meet no resistance.

Ethical Analysis: This scenario represents multiple violations of International Humanitarian Law, including the principles of distinction (discriminating between combatants and civilians), proportionality, and the prohibition on indiscriminate attacks. It illustrates how autonomous systems can be deliberately calibrated to commit what would legally constitute war crimes, with the veneer of technical precision masking fundamentally unlawful targeting criteria. The delegation of life-and-death decisions to algorithms, combined with biased “threat scoring” that treats civilians as presumptive threats, exemplifies the accountability gap and moral disengagement discussed in Section II of this document.

References

[1] Tekna, “Autonome våpen og etikk,” Professional development course, 2024, verified: Tekna professional development course on autonomous weapons and ethics. [Online]. Available: https://www.tekna.no/kurs/autonome-vapen-og-etikk-50926/

[2] International Committee of the Red Cross, “A legal perspective: Autonomous weapon systems under international humanitarian law,” 2024, iCRC position paper. [Online]. Available: https://www.icrc.org/en/document/autonomous-weapon-systems-under-international-humanitarian-law

[3] R. A. Seehuus, “Foredrag: Autonomi i militære operasjoner,” 2024, nORDSEC foredrag og FFI forskning på dronesvermer. [Online]. Available: https://www.nordsec-cluster.no/aktuelt/foredrag-autonomi-i-milit%C3%A6re-operasjoner

[4] United Nations Secretary-General, “Lethal autonomous weapons systems,” 2024, report of the Secretary-General. [Online]. Available: https://docs.un.org/en/A/79/88

[5] A. Lysbakken, L. Haltbrekken, N. Wilkinson, and P. Eide, “Representantforslag om at norge må innta en lederrolle i kampen mot dødelige autonome våpensystemer (drapsroboter),” 2021, representantforslag fra SV. [Online]. Available: https://www.stortinget.no/no/Saker-og-publikasjoner/Saker/Sak/?p=84405

[6] Forsvarsdepartementet, “Forsvarssektorens ki-strategi,” 2024, norwegian Ministry of Defence AI strategy. [Online]. Available: https://www.regjeringen.no/no/dokumenter/forsvarssektorens-ki-strategi/id3057089/

[7] R. Crootof and D. Richemond-Barak, “The future of warfare: National positions on the governance of lethal autonomous weapons systems,” 2024, lieber Institute policy brief. [Online]. Available: https://lieber.westpoint.edu/future-warfare-national-positions-governance-lethal-autonomous-weapons-systems/

[8] R. Crootof, “What is meaningful human control, anyway? cracking the code on autonomous weapons and human judgment,” 2024, analysis of meaningful human control concept. [Online]. Available: https://mwi.westpoint.edu/what-is-meaningful-human-control-anyway-cracking-the-code-on-autonomous-weapo

[9] H. M. Roff, “The strategic robot problem: Lethal autonomous weapons in war,” *Journal of Military Ethics*, vol. 15, no. 1, pp. 37–54, 2016. [Online]. Available: https://doi.org/10.1080/15027570.2016.1174562

[10] J. R. Boyd, “The essence of winning and losing,” *Air University Review*, vol. 37, no. 4, pp. 2–11, 1986, oODA Loop concept development. [Online]. Available: https://www.airuniversity.af.edu/Portals/10/AUPress/Books/B_0151_BOYD_DISCOURSE_WINNING_LOSING.PDF

[11] North Atlantic Treaty Organization, “Nato artificial intelligence strategy,” Brussels, 2021, nATO official strategy document. [Online]. Available: https://www.nato.int/cps/en/natohq/official_texts_187617.htm

[12] T. Lyshaug, “Hva skal vi med autonome våpen: En litteraturstudie om autonome våpensystemer,” Master thesis, Forsvarets høgskole, 2021, norwegian Defence University College master’s thesis - comprehensive literature review on autonomous weapons systems. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/2835213

[13] D. Fudenberg and J. Tirole, *Game Theory*. Cambridge, MA: MIT Press, 1991, standard game theory textbook covering Nash equilibria, repeated games, and cooperation.

[14] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge, UK: Cambridge University Press, 1990, nobel Prize-winning analysis of collective action problems and enforcement mechanisms.

[15] D. D. Heckathorn, “Collective action and the second-order free-rider problem,” *Rationality and Society*, vol. 1, no. 1, pp. 78–100, 1989, analysis of enforcement challenges in collective sanction systems.

[16] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, “3d gaussian splatting for real-time radiance field rendering,” in *ACM Transactions on Graphics*, vol. 42, no. 4, July 2023, verified: arXiv accessible, ACM DL DOI accessible - SIGGRAPH 2023 Best Paper Award. [Online]. Available: https://arxiv.org/abs/2308.04079

[17] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, “Segment anything,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2023, verified: arXiv accessible, ICCV 2023 - Meta AI foundational segmentation model. [Online]. Available: https://arxiv.org/abs/2304.02643

[18] R. A. Seehuus and A. Pettersen, “Ffi har utviklet en sverm med angrepsdroner,” 2024, fFI drone swarm technology development. [Online]. Available: https://www.ffi.no/aktuelt/nyheter/ffi-har-utviklet-en-sverm-med-angrepsdroner

[19] C. v. Clausewitz, *On War*, M. Howard and P. Paret, Eds. Princeton, NJ: Princeton University Press, 1984, classic military theory text introducing center of gravity concept in Books VI and VIII.

[20] European Union, “Artificial intelligence act,” 2024, eU Regulation on AI systems. [Online]. Available: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32024R1689

[21] Y. Abraham, +972 Magazine, and Local Call, “’lavender’: The ai machine directing israel’s bombing spree in gaza,” April 3 2024, verified: Investigative report by +972 Magazine and Local Call based on interviews with six Israeli intelligence officers - April 2024. [Online]. Available: https://www.972mag.com/lavender-ai-israeli-army-gaza/

[22] Human Rights Watch, “Questions and answers: Israeli military’s use of digital tools in gaza,” September 10 2024, verified: HRW analysis of four AI targeting systems - September 2024. [Online]. Available: https://www.hrw.org/news/2024/09/10/questions-and-answers-israeli-militarys-use-digital-tools-gaza

[23] Office of the High Commissioner for Human Rights, “Gaza: Un experts deplore use of purported ai to commit ’domicide’ in gaza,” April 2024, verified: UN experts statement on AI systems Gospel, Lavender, and Where’s Daddy - April 2024. [Online]. Available: https://www.ohchr.org/en/press-releases/2024/04/gaza-un-experts-deplore-use-purported-ai-commit-domicide-gaza-call

[24] M. N. Schmitt, “The gospel, lavender, and the law of armed conflict,” 2024, verified: Legal analysis of IDF AI systems under LOAC - Lieber Institute 2024. [Online]. Available: https://lieber.westpoint.edu/gospel-lavender-law-armed-conflict/

[25] A. Skøghøy and H. M. Lomell, “Resistance to platformization: Palantir in the norwegian police,” *Information, Communication & Society*, 2024, verified: Academic analysis of failed Palantir Omnia project in Norwegian police 2016-2018. [Online]. Available: https://www.tandfonline.com/doi/full/10.1080/1369118X.2024.2325533

[26] NordForsk CUPP Research Consortium, “Critical understanding of predictive policing,” 2024, verified: Nordic research project on algorithmic policing in Denmark, Estonia, Latvia, Norway, Sweden and UK - 2021-2024. [Online]. Available: https://www.nordforsk.org/projects/critical-understanding-predictive-policing

[27] K. Mikkelsen, “Implementing lethal autonomous weapon systems into mission command leadership,” Bachelor thesis, Luftkrigsskolen, 2023, norwegian Air Force Academy bachelor thesis on LAWS integration. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/3084073

[28] T. Haugen, “Effektiv kommando og kontroll av f-35: En vurdering av autonomi sin rolle innenfor f-35 operasjoner,” Master thesis, Forsvarets høgskole, 2022, norwegian Defence University College master’s thesis on autonomy in F-35 operations. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/3021251

[29] S. Kvam, “Autonom målutvelgelse: Et rettslig dilemma?” Master thesis, Forsvarets høgskole, 2017, norwegian Defence University College master’s thesis on legal implications of autonomous target selection. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/2477938

[30] A. Q. Ekren, “Hærens evne til digitalisering: Et organisasjonskulturelt perspektiv,” Master thesis, Forsvarets høgskole, 2024, norwegian Defence University College master’s thesis on Army digitalization capabilities. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/3179416

[31] H. Øie, “Ubemannede maritime systemer: Fremtidens styrkemultiplikator, eller juridisk hodebry!” Master thesis, Forsvarets høgskole, 2015, norwegian Defence University College master’s thesis on unmanned maritime systems and international law.

[32] J. Stomperudhaugen, “Droner i sjøforsvaret: Muligheter og begrensninger for oppdragsløsning i undervannsdomenet,” Master thesis, Forsvarets høgskole, 2025, norwegian Defence University College master’s thesis on naval drone applications. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/3204844

[33] M. Riis Asdahl, J. U. Riis, and P. A. N. Wandas, “Autonom kollisjonsunnvikelse på maritim dronesverm,” 2023, norwegian Naval Academy bachelor thesis on autonomous collision avoidance for maritime drone swarms.

[34] M. S. Lund, “Autonomi i cyberoperasjonar,” in *Cyberkrigsføring*. Høgskolen i Innlandet/Forsvarets høgskole, 2023, chapter on autonomy in cyber operations. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/3130350

[35] H. Syse, *Ethics, War and International Law*. London: Palgrave Macmillan, 2018.

[36] Luftled, “En sverm av smarte samarbeidende droner,” 2024, norwegian defense technology analysis. [Online]. Available: https://luftled.info/en-sverm-av-smarte-samarbeidende-droner/

[37] T. Haugen, “Studie av autonome våpensystemer,” 2022, norwegian Defence University College thesis on autonomous weapons systems. [Online]. Available: https://fhs.brage.unit.no/fhs-xmlui/bitstream/handle/11250/3021251/%5b29%5d%20Trond%20Haugen_OPG5101_Trond%20Haugen_2022.pdf?sequence=1&isAllowed=y

[38] Forsvarets høgskole, “Fremtidens autonome ubemannede kapasiteter i sjøforsvaret,” 2018, study on future autonomous unmanned capabilities in the Norwegian Navy. [Online]. Available: https://fhs.brage.unit.no/fhs-xmlui/bitstream/handle/11250/2568400/Fremtidens%20autonome%20ubemannede%20kapasiteter%20i%20Sj\T1\oforsvaret.pdf?sequence=1&isAllowed=y

[39] E. P. M. Ponta, “Drone versus tjenestehund: Substitusjon eller komplementære kapabiliteter?” Master thesis, Forsvarets høgskole, 2024, norwegian Defence University College master’s thesis comparing drones and service dogs.

[40] O. S. Hareide, T. Relling, A. Pettersen, A. Sauter, F. V. Mjelde, and R. Ostnes, “Fremtidens autonome ubemannede kapasiteter i sjøforsvaret,” 2018, norwegian Naval Academy study on future autonomous unmanned capabilities. [Online]. Available: http://fhs.brage.unit.no/fhs-xmlui/handle/11250/2568400