The human firewall: Preserving military decision-making advantage in the age of autonomous systems and agentic AI

Author: *Mark Baines*

Introduction

Agentic AI and autonomous systems are moving from experimental technology into mature defence and security capabilities. They promise faster sensing, analysis, planning, targeting, logistics, cyber defence, and decision support. But in military and political contexts, the central issue is not productivity. It is legitimacy of command and control.

Defence and national security leaders must exploit the operational advantage of AI without allowing machine speed to erode human judgement, democratic accountability, alliance cohesion, or compliance with law and ethics. The core task is to design human-machine systems in which AI accelerates understanding and execution, while commanders retain authority over intent, escalation, proportionality, and consequence.

This paper argues that the human advantage in the context of military operations is the ability to exercise strategic judgement and human agency under political constraint. Autonomous systems can compress decision cycles, but only humans can weigh legitimacy, alliance politics, national interest, collateral damage and escalation risks. In the age of algorithmic warfare, the operational advantage will rest of the ability to combine machine tempo with human command.

The evolving speed of war

Stripped of its political and societal window dressing, war remains the application of state-sanctioned violence to achieve an outcome in the national interest. Clausewitz warned that "[i]t would be futile – even wrong – to try and shut one's eyes to what war really is from sheer distress at its brutality." This remains as true today as when those were words were written 200 years ago. The nature of war remains violent and distressing, but the speed of war, now augmented by AI, is changing and our human role in conducting it is evolving.

The strategic conversation about AI is often framed as a competition for technological advantage: which state can deploy autonomy fastest, integrate data most effectively, and shorten the observe–orient–decide–act loop. Those questions matter. However, in defence and national security, the more consequential question is whether AI will augment and improve human command or gradually displace the judgement that gives command its legitimacy.

Autonomous systems are changing the economics and tempo of conflict. They can help forces see more, decide faster, distribute risk, protect personnel and operate at a scale that human-only formations cannot match. Yet the source of strategic advantage will not be autonomy alone. It will be the discipline to decide where machines should accelerate operations and when humans must retain control, and how political leaders, commanders and institutions preserve accountability under pressure.

The military strategic context: Speed, saturation, and political risk

The return of large-scale war in Europe, pervasive grey-zone competition, cyber operations, information warfare, drone proliferation, and renewed great-power rivalry have made AI and autonomy strategically unavoidable. Militaries are seeking systems that can process vast sensor feeds, support operational planning, defend networks, coordinate unmanned platforms, and enable faster decisions across multiple domains.

A further strategic driver is the asymmetric balance of investment. Recent conflicts show that relatively low-cost drones, loitering munitions, and decoys can force defenders to expend far more expensive interceptors, aircraft sorties, sensors, and command attention. The economic exchange can favour the attacker even when the defender is tactically successful, because each intercepted drone may still consume scarce munitions, reveal defensive positions, or impose disproportionate financial, social, and industrial costs.

This creates a strategic investment dilemma for advanced militaries. Forces built around exquisite, expensive platforms and limited inventories can be placed under pressure by adversaries able to manufacture, adapt, and deploy cheaper expendable systems at scale. The answer is to rebalance investment towards layered defence, resilient supply chains, and procurement models that can match the tempo of battlefield evolution.

For leaders, cost asymmetry turns AI and autonomy into a fiscal and industrial strategy challenge as much as an operational one. The decisive question is whether defence investment is properly aligned to the new economics of conflict: enough high-end capability to deter and defeat sophisticated threats, enough low-cost autonomous mass to absorb attrition and impose costs, and enough adaptive industrial capacity to replenish, upgrade, and learn faster than the adversary.

But military AI is more than another productivity tool. In conflict, speed changes politics. Faster decision cycles can improve survivability and deterrence, but they can also compress deliberation, increase dependence on opaque systems, mask alternatives, magnify automation bias, and reduce the time available for judgement. The danger is that democratic states adopt AI in ways that weaken the legal accountability and strategic restraint that undermines the moral character of the state and tilts towards the dystopian application of automated violence.

This reframes the agenda for leaders. AI must enable decision superiority while preserving political control, operational accountability, interoperability with allies, and public trust.

AI is already changing warfare

The war in Ukraine has become the clearest contemporary example of how AI is entering drone warfare, not as science fiction autonomy at scale, but as a set of practical capabilities that help unmanned systems survive electronic warfare, find targets, prioritise information and support faster operational decisions.

Machine-speed intelligence triage. Drones generate huge volumes of imagery and sensor data. AI and machine learning can help detect movement, classify objects, highlight anomalies, and cue human operators to items of interest. It changes where commanders' judgement is applied, shifting people from raw observation to verification, prioritisation, and decision.

AI-assisted target recognition and terminal guidance. Ukrainian units have used drones with AI-enabled targeting to help lock onto vehicles and logistics assets even when communications links are degraded. The strategic significance is not that the machine replaces the commander, but that AI can help maintain effectiveness in the final phase of a strike when jamming, distance, or poor connectivity would otherwise degrade human control.

GPS-denied navigation under electronic attack. Heavy jamming and spoofing have pushed Ukrainian and Russian developers towards drones that rely less on satellite navigation and continuous radio control. AI-enabled computer vision and onboard processing are being used to help drones interpret terrain and maintain navigation, continuing missions when GPS or uplinks are unreliable.

AI-enabled logistics interdiction. Ukraine's mid-range drone campaign illustrates how AI can support strikes against logistics networks deep behind the front. The military effect is cumulative: degrading supply chains can reduce artillery fire and slow movement, making it harder for forces to mass, without relying solely on traditional deep-strike assets.

Counter-drone adaptation. The drone war is also an AI-enabled contest between attack and defence. As Russia has adapted Shahed/Geran-style drones, Ukraine has developed layered countermeasures, including electronic warfare, interceptors, and detection systems. This points to a broader pattern. AI does not produce a permanent edge, but accelerates

the age-old cycle of adaptation between offence, defence, and counter-countermeasure.

The leadership lesson here is clear. AI in drone warfare is currently less about fully independent machines making strategic choices and more about narrow autonomy at the edge: navigation, recognition, tracking, prioritisation, and resilience under electronic attack. These capabilities can create operational advantage, but they also intensify the need for clear human authorisation, target validation, and auditability.

Ethical challenges: Machine speed meets human consequence

The ethical challenge in military AI is not limited to fully autonomous weapons. It begins earlier, with the systems that sort intelligence, recommend targets, prioritise threats, classify behaviour, and shape what commanders see. Even when a human remains formally "in the loop," AI could alter the decision environment so profoundly that human control becomes nominal rather than meaningful.

Meaningful human control. A commander may technically approve a strike, but if the recommendation arrives under time pressure, with limited explainability and a high confidence score, the human role can become a rubber stamp. The ethical question is whether the person who clicked "Approve" had enough information, time, authority, and situational understanding to exercise genuine judgement. Human analysts and commanders make errors. Is human error more acceptable and less likely than algorithmic error?

Target misclassification and civilian harm. AI systems trained on imperfect or context-poor data may misread civilian behaviour as hostile, confuse civilian objects with military assets, or fail to recognise protected people and places. In dense urban environments, the ethical stakes are acute because small classification errors can produce irreversible harm.

Automation bias. Operators and commanders may place excessive trust in machine outputs because they appear objective, precise, and definitive. This can suppress dissenting judgement and narrow the range of options considered. It can make it harder for humans to challenge a system that appears confident but may be wrong.

Accountability gaps. When a harmful outcome results from a chain involving commanders, operators, software developers, data providers, contractors, and allied systems, responsibility can become diffuse. The harder it is to explain why an AI-enabled system produced a recommendation or action, the harder it becomes to assign responsibility, learn lessons, and maintain public legitimacy.

Data bias and digital dehumanisation. AI systems may encode patterns from historical data that reflect surveillance gaps, operational bias, or discriminatory assumptions. In military settings, this can turn people into data signatures and target profiles, weakening the moral friction that should accompany decisions about force.

Risk management versus risk avoidance. Political and military leadership that becomes overly confident in machine outputs may unwittingly accept higher risk without fully understanding the consequences should that risk be realised. The war in Iran is a case in point. There is already a political and military cultural bias to be seen as active risk managers, not risk avoiders, that is likely to be reinforced as AI systems become more pervasive. Framing risk around the question of consequence acceptance needs to remain a human centred activity to maintain accountability and responsibility.

Escalation and loss of strategic control. AI-enabled systems can accelerate action-reaction cycles between adversaries. If each side relies on faster automated detection, warning, and response systems, crises may escalate before political leaders have time to interpret intent, communicate restraint, or correct false signals.

Military AI must therefore be assessed not only for performance, but for explainability, predictability, data provenance, bias, auditability, human override, legal review, and the quality of human judgement it enables under pressure.

AI risks in intelligence analysis

Intelligence analysis is one of the areas in which AI can deliver major gains: faster collection triage, cross-source discovery, translation, anomaly detection, pattern recognition, and briefing support. But it is also one of the areas where AI failure can be strategically dangerous, because intelligence products shape warning, targeting, and crisis decision-making.

Adversarial data and deception. Intelligence systems are exposed to deliberate manipulation: spoofed imagery, synthetic media, false social signals, poisoned datasets, and adversary-controlled information streams. AI may identify patterns in this data without recognising that the pattern has been seeded to mislead.

Hallucinated certainty. Generative systems can produce fluent assessments that contain fabricated details, unsupported links, or misleading confidence. In intelligence settings, in addition to factual error, the risk it is that a plausible but unverified narrative becomes embedded in a briefing, decision paper, or operational plan.

Weak source provenance. AI can blend classified reporting, open-source material, social media, commercial data, and previous analysis into a seamless answer. Unless provenance is explicit, analysts and commanders may lose sight of what is confirmed, what is inferred, what is single-source, and what comes from lower-quality or potentially compromised reporting.

Automation bias in assessment. Analysts may over-weight AI-generated judgements because they appear comprehensive, data-driven, or neutral. This can suppress dissent and weaken structured analytic techniques, reducing decision makers' willingness to challenge a machine-generated consensus.

Analytic monoculture. If multiple agencies or coalition partners rely on similar models, datasets, or vendor platforms, they may converge on the same errors. AI can create the illusion of independent corroboration when different outputs are drawing from overlapping inputs, assumptions, or training data.

Loss of analytic tradecraft. If analysts become editors of AI-generated products rather than investigators of evidence, organisations risk weakening the habits that make intelligence valuable: source scepticism, alternative hypotheses, confidence calibration, red teaming, and awareness of deception.

Analysis of alternatives. AI blended analysis can mask alternative interpretations, limiting the range of possible responses and courses of action. Just as confidence can be misplaced on AI generated judgements, confidence in the weighting AI uses to discard, ignore or value intelligence can be equally as dangerous, especially as it is likely to be more insidious and less well understood or interrogated by human analysts.

AI should be used to accelerate exploitation and test hypotheses, while human analysts retain responsibility for source evaluation, confidence judgements, and alternative explanations. Every AI-assisted intelligence product should make clear what the system used, what it inferred, what it could not verify, and where human judgement remains decisive.

Human advantage will remain paramount

The enduring human advantage in military and political decision-making is not speed of calculation. It is agency: strategic judgement under constraint, the capacity to understand purpose, interpret intent, weigh proportionality, manage escalation, sustain alliances, maintain legitimacy, and take responsibility for consequences.

AI can identify a pattern in sensor data. It cannot determine the political meaning of acting on it. It can recommend a course of action. It cannot judge whether that action aligns with national objectives, alliance commitments, or the laws of armed conflict. It can accelerate targeting workflows. It cannot bear moral responsibility for civilian harm or decide what risks a democracy should accept in pursuit of security.

This distinction matters because the most consequential military operational decisions are rarely clear cut; often routinely confounded by "the fog of war". These decisions happen amid uncertainty, deception, and incomplete or conflicting intelligence. They require commanders and political leaders to understand not only what can be done, but what should be done, and what a decision will communicate to adversaries and allies.

Consider the difference in practise. In a drone-saturated battlespace, AI can detect movement, classify objects and prioritise threats. Commanders must decide whether the context justifies action. In cyber defence, AI can identify anomalous behaviour and trigger containment. Leaders must judge when a technical incident becomes a strategic signal. In coalition operations, AI can fuse intelligence across partners. Humans must manage classification, trust and political caveats.

The advantages that will endure

The most durable human advantages in the military and political domain are those rooted in legitimacy, responsibility, and context. They include command judgement, ethical reasoning, political sensitivity, escalation awareness, mission command, empathy for civilian consequences, and the ability to interpret adversary intent beyond data patterns. Added to this are leadership in adversity and the determination and flair to defeat fanatics.

These are not soft constraints on hard power. They are part of hard power. Military effectiveness in democratic societies depends on legal authority, public consent, coalition legitimacy, and the ability to align tactical action with strategic purpose.

The least durable human advantages are those based on routine information processing: manual intelligence triage, repetitive reporting, standard logistics optimisation, basic operational planning support, document discovery and pattern recognition. These functions will increasingly be augmented or automated. The premium will shift to leaders who can decide which patterns matter, which outputs to distrust, which risks require human intervention, and which decisions cannot be delegated.

This distinction should shape force design, capability development, and civil-military education. Armed forces will need more people who can operate with machines, question and command through "smart" machines, without becoming subordinate to them.

The implications for command

The best military use of AI is to ask, "Which decisions require human command, and which tasks can be delegated to machines so humans have more time, clarity and resilience for those decisions?"

AI should accelerate sensing, synthesis, and option generation, not replace command intent. It should widen the decision space, not narrow it to the most machine-readable answer. It should expose assumptions, not bury them behind confidence scores. Used well, autonomous systems become a force multiplier: they extend reach, tempo, and resilience while leaving humans responsible for purpose, authorisation, and consequence.

This requires explicit design. Commanders need to know which functions are advisory, which are supervised, which are bounded for autonomous execution, and which remain prohibited without human authorisation. Political leaders need assurance that operational tempo does not outrun policy control. Alliances need common standards so AI-enabled forces can fight together without creating unmanaged risk.

A useful rule is this: Delegate speed, scale and sensing to AI. Retain intent, proportionality, and escalation control under humans.

AI can prepare the operational picture, but commanders must decide what the mission needs. AI can recommend a response to a cyber intrusion, but leaders must judge whether it is law enforcement, espionage, coercion or armed attack. AI can optimise resupply through contested terrain, but commanders must apply the principles of war and weigh mission priorities against risk to life. AI can analyse influence operations, but political leaders must decide how a democracy responds without undermining its own values.

Attributes of commanders in the age of algorithms and autonomy

Command in the age of algorithms will require more than technical familiarity with AI-enabled systems. It will require a new blend of military judgement, digital literacy, ethical courage and political awareness. The commander's role will need to deliberately grow to include orchestrating human-machine teams under conditions of uncertainty, speed and contested information. (See Table 1.)

Table 1 | Attributes of commanders

Attribute	Why it matters
Algorithmic literacy	Commanders must understand what AI systems can and cannot do, including their dependence on data quality, training context, uncertainty, model drift, and vulnerability to deception. Models that share the same algorithmic underpinning will share the same strengths and weaknesses.
Judgement under machine-speed pressure	AI will compress decision timelines. Commanders must be able to generate the head space to properly consider decisions when consequences are strategic, legal or ethical, even when systems recommend rapid action.
Healthy scepticism	Commanders need the confidence to challenge confident machine outputs, ask what data is missing, test assumptions, look for alternatives, and resist automation bias.
Mission command in human-machine teams	Commanders will need to express intent clearly enough for people, software and autonomous systems to operate coherently within authorised boundaries.
Ethical courage	Commanders must be willing to pause, override or reject AI-enabled recommendations when legality, proportionality, distinction, escalation risk or legitimacy is uncertain.
Systems thinking	AI-enabled operations depend on data, sensors, networks, cyber resilience, logistics, doctrine, training and governance. Commanders must understand the system, not only the platform.
Political and strategic awareness	In algorithmic warfare, tactical actions can rapidly become strategic signals. Commanders must understand how AI-enabled effects may be interpreted by allies, adversaries, publics and political leaders.
Learning agility	AI-enabled conflict will reward forces that adapt quickly. Commanders must be comfortable with rapid experimentation, feedback loops, red-teaming, and continuous updating of doctrine and practice.
Accountability ownership	The presence of AI does not dilute command responsibility. Commanders must remain accountable for how systems are authorised, supervised, challenged, and used. They must understand and be prepared to accept consequences.

Shifting from prompt engineering to command engineering

The fashionable civilian skill is prompt engineering. In defence, the emergent discipline is command engineering: designing the human-machine conditions under which lawful, ethical, and strategically coherent decisions are made at speed.

Command engineering means specifying decision rights, authorisation thresholds, data provenance, confidence requirements, override mechanisms, audit trails, red-team testing, fail-safe modes, and escalation triggers. It also means training commanders and ministers to interrogate AI-enabled recommendations. The best commanders will not be those who either distrust AI or defer to it. Understanding that intelligence is not the same as knowledge. The best commanders will be those who know when to exploit machine speed and when to demand human deliberation.

Next-level complexity: AI and coalition operations

AI will have a particularly significant effect on coalition-of-the-willing operations, where forces may share a mission but not identical doctrine, systems, data rules, political mandates, or risk appetites. In these settings, AI can strengthen coalition effectiveness by creating shared understanding and accelerating coordination. But it can also expose fractures in trust, interoperability, and accountability.

This is no longer theoretical. Platforms such as Palantir's Maven Smart System, derived from the broader Project Maven ecosystem, are designed to fuse data from satellites, drones, geospatial intelligence, sensors, and operational reporting to support battlefield awareness, target development, and command decision-making. Anduril's Lattice similarly illustrates the shift towards AI-enabled command-and-control architectures that integrate sensors, autonomous systems, and counter-drone responses into a common operating environment. These tools show how AI is moving from analytic support into the operational nervous system of modern coalitions.

Shared operational pictures. AI can help fuse sensor feeds, intelligence reports, open-source information, satellite imagery, cyber indicators, and partner reporting into a coalition common operating picture. The leadership benefit is faster shared awareness. The challenge is that different partners may not trust the same data sources, may classify information differently, or may be unwilling to expose sensitive collection methods.

Coalition targeting and approval chains. AI-enabled targeting support can help identify patterns, prioritise threats and accelerate collateral damage estimation. In a coalition of the willing, however, each nation may apply different legal interpretations, rules of engagement, or political caveats. A machine-generated recommendation may be technically persuasive but politically unusable for some partners.

Interoperability between unequal digital forces. AI can help translate between systems, formats, and operating pictures, but it cannot fully compensate for incompatible architectures, uneven data quality, or partners with different levels of digital maturity. In practice, the best-equipped force may set the AI-enabled tempo, while others struggle to keep pace or become dependent on a lead nation's systems.

Political caveats encoded into operations. Coalition operations often depend on national caveats: where forces can operate, what effects they can support, which data they can share, and what risks they can accept. AI-enabled command systems will need to make these constraints visible and enforceable rather than treating the coalition as a single unified actor.

Assurance and liability across partners. If an AI-enabled recommendation contributes to a harmful outcome, responsibility may be distributed across national commanders, system providers, data owners, and coalition headquarters. Coalition leaders will need pre-agreed assurance standards, audit trails, and incident-review mechanisms before crises occur, not after failure.

The relevance of systems such as Maven and Lattice is not that any one vendor will define the future battlefield. It is that intelligence, targeting and command tools are becoming digital ecosystems: they ingest diverse data, apply machine learning, connect decision-makers and enable rapid integration of third-party capabilities. In coalition settings, this raises difficult questions about who controls the data layer, whose algorithms shape the operating picture, and how national caveats are respected inside shared digital infrastructure.

The leadership implication is that coalition AI is as much a diplomatic and governance challenge as a technical one. Coalitions will need shared data standards, federated architectures, zero trust networks, agreed confidence thresholds, human-control principles, and political processes for resolving disagreement when machine-speed recommendations collide with national caveats.

Conclusions

The future will belong to those commanders that can combine machine-speed capability with human political judgement and disciplined military leadership, exercised with flair.

The broader risk is that human institutions become too passive: accepting machine recommendations because they are fast and trusting confidence scores because they create the illusion of precision. To maximise human advantage, we should:

- **Protect command competence.** Ensure ministers, officials, and officers still learn to reason, challenge intelligence, understand adversaries, and practise making decisions under uncertainty rather than merely supervising machine outputs.
- **Reframe AI strategy around decision superiority, not automation.** Track whether AI improves operational understanding, speed of responsible action, resilience, deterrence, and strategic coherence, not just process efficiency.
- **Map autonomy by consequence.** Distinguish between low-risk automation, supervised decision support, bounded autonomous action, and decisions that require explicit human authority.

In the age of military AI, ethics cannot sit outside strategy as a review process or compliance overlay. It is part of command itself. Leaders who cannot explain how an AI-enabled capability preserves human judgement, legal accountability, and political control will struggle to sustain legitimacy when the system is tested in crisis.

This makes ethical leadership a source of operational and strategic advantage. Forces that build trust into their AI systems – through clear authority, disciplined oversight, transparent assurance, and meaningful human control – will be better able to move fast without losing coherence. They will also be better placed to work with allies, maintain public confidence, and distinguish lawful military power from automated violence.

The leadership task is to integrate innovation and ethics. The military organisations that succeed will be those that treat ethical design, accountable command, and human judgement as the architecture that allows AI to be used with confidence at speed.

The defining strategic question of the AI era is "How do we preserve decision advantage when machines can sense, decide and act faster than the institutions that authorise them?" The answer will determine not only operational advantage, but the legitimacy of democratic power in an age of algorithmic warfare.

Get in touch to discuss how defence and national security leaders can harness AI and autonomous systems while preserving human judgement, accountable command, and strategic control.

Listen to Mark Baines discuss the issues raised in this paper here.

Connect with Mark Baines on LinkedIn.