Thrive Or Die - UK Defence And Aerospace

UK Defence And Aerospace Edition — Where The Gap Stops Being An Analogy

A Charterhouse Consultants white paper by Bridget Convey, October 2025

Executive Summary

Every other paper in this series treats the Bengio and Hinton warning about autonomous systems acting with reduced human oversight as a mechanism that explains corporate risk. In UK defence, it is not an analogy. It is the literal, live policy question. In June 2026, a UK armed forces minister stated publicly that Britain must prepare for taking the human out of the loop in weapons systems, directly contradicting the Ministry of Defence's own stated doctrine that there must be context-appropriate human involvement in any weapon that identifies, selects and attacks a target. That contradiction, inside the same government, in the same year, is the clearest evidence in this entire series that the capability to governance gap is real, urgent, and already being argued over at the highest level.

The Premise: The Gap Is The Danger, Not The AI

In October 2023, a group including Turing Award winners Yoshua Bengio and Geoffrey Hinton, alongside Stuart Russell, Daniel Kahneman and more than twenty other senior AI researchers, published a consensus paper in the journal Science on the risks of advanced AI. Their argument was precise. AI capability and autonomy are increasing rapidly. Companies and institutions are shifting toward systems that can act and pursue goals with reduced human oversight. Governance is not keeping pace with that shift. The danger they describe does not come from AI itself. It comes from the widening gap between what a system can now do on its own and how closely anyone is still watching it.

That paper is about safety and governance at a societal scale. It does not comment on corporate or institutional strategy, and I want to be precise about that rather than overstate its relevance. What it gives this paper is not an endorsement, it gives us the mechanism. Capability grows. Autonomy grows. Oversight does not grow at the same rate. In UK defence specifically, research modelling AI behaviour in simulated crises found AI systems chose escalatory, rather than de-escalatory, action in a striking majority of scenarios, a finding directly relevant to a government now doubling its investment in autonomous platforms while its own doctrine and its own ministers appear to disagree on how much human control should remain.

This is the basis for the claim in the title. Thrive or die is not a slogan layered on top of the argument. It is the plain consequence of the mechanism. An organisation that builds governance at the same pace it builds AI capability closes the gap and survives it. One that lets capability outrun governance does not get a soft landing.

Thrive Or Die: The Pattern In Organisations That Did Not Close The Gap

Kodak invented the first digital camera in 1975 and its own internal research predicted by 1981 that digital would eventually replace film. What it did not close was the governance gap between that capability and a business model still built entirely around film margins. The company filed for bankruptcy in 2012, thirty seven years later. Blockbuster was offered the chance to buy Netflix outright for fifty million dollars in 2000, declined, and filed for bankruptcy in 2010, having eliminated close to eighty four thousand jobs in the collapse. Both prove the same point. The danger was never a lack of capability. It was a governance gap left open until it was unmanageable.

Kodak and Blockbuster are useful because they show the cost of a slow, commercial gap. Defence shows what the same gap looks like when the stakes are immediate and irreversible rather than financial. The UK Ministry of Defence's stated position, since its 2022 Defence AI Strategy, is that Human-Machine Teaming, not full autonomy, sits at the heart of its approach, and that there must be context-appropriate human involvement in any weapon system that identifies, selects and attacks a target. A House of Lords committee has formally urged the government to proceed with caution and ensure proper democratic oversight of autonomous weapons development. Yet in June 2026, the parliamentary under-secretary of state for the armed forces argued publicly that Britain must prepare for the possibility of taking the human out of the loop, on the grounds that adversaries would not hesitate to deploy systems that kill without human approval. This is not a hypothetical future scenario from this paper. It is a live, unresolved contradiction inside UK government policy today, and it is exactly the capability-outrunning-governance mechanism the Science consensus paper describes, playing out at the point where the consequences are most severe.

The Evidence In UK Defence: Investment Accelerating, Doctrine Unresolved

Two facts anchor this paper. Investment in autonomous capability is accelerating sharply. The governance question of how much human control should remain is, by the government's own public statements, still unresolved.

Investment Doubling While The Doctrine Question Stays Open

The UK government has doubled its funding commitment for autonomous systems from £2 billion to £4 billion over the current parliament, intended to support the British Army's stated objective of a tenfold increase in lethality over the next decade. The Ministry of Defence has awarded more than two dozen contracts for AI targeting systems, alongside successful trials of AI-powered drone technology for landmine and explosive ordnance detection. This is capability accelerating in exactly the way the Science consensus paper describes, deliberately, well-funded, and fast. The open question is whether the governance structure around that capability, currently stated as Human-Machine Teaming with context-appropriate human involvement, is scaling at the same pace as the £4 billion now behind it.

The Escalation Research: What Happens When Oversight Is Thin

Independent research modelling AI decision-making in simulated international crises found that AI systems chose escalatory action, rather than restraint, in a striking majority of scenarios, a finding that directly supports the mechanism this whole paper series is

built on. Capability without matching oversight does not default to the cautious option. It defaults to the option the system was optimised for, which is not necessarily the safe one. A House of Lords select committee report, Proceed With Caution, has formally recommended the government incorporate a specific, defined concept of meaningful human control into its Defence AI policy, and continue developing that concept as the technology changes, rather than leaving the current, more general 'context-appropriate' standard to carry the weight of every future decision.

UK Defence And Aerospace In Three Horizons: The Gap Closes Or It Does Not

If the gap between capability and governance is the real variable, then each horizon below has two outcomes, not one. These are deliberately stated as firm positions, not hedged possibilities, because a board conversation needs something to agree or disagree with.

Two Years: 2026 To 2028

What happens everywhere: autonomous targeting and decision-support systems move from trial to operational deployment across UK defence procurement.

- Thrive path. The government resolves the current doctrinal contradiction by adopting the House of Lords committee's recommendation, a specific, legally defined standard for meaningful human control, rather than leaving 'context-appropriate involvement' to be interpreted differently by different parts of government.
- Die path. Investment in autonomous capability continues to outpace resolution of the human control question, and the doctrine remains whatever the most recent minister's public statement implies, rather than a settled, auditable standard.

Five Years: 2028 To 2031

What happens everywhere: UK defence AI systems reach the scale where an incident involving reduced human oversight becomes a realistic possibility, not a hypothetical one.

- Thrive path. The UK has, by this point, built the technical assurance and testing standards the House of Lords committee called for, including data quality, human-machine interaction, and transparency standards specific to autonomous weapons, giving government and industry a shared, defensible standard to operate against.
- Die path. The UK has continued procuring and deploying autonomous capability faster than it has built assurance standards, leaving the country exposed to the same reputational, legal and human cost Kodak and Blockbuster faced commercially, but in a domain where the cost of a governance failure is not measured in jobs or market share.

Ten Years: 2031 To 2036

What happens everywhere: the international norm on autonomous weapons, currently unresolved, either converges around meaningful human control standards or fragments as different nations adopt different thresholds.

- Thrive path. The UK's early investment in assurance and testing standards, rather than only in capability, becomes an exportable framework other nations adopt, in the way the UK has historically exported financial and legal regulatory standards.
- Die path. The UK's doctrine remains reactive to whichever adversary capability prompts the next policy statement, rather than a settled position, leaving both the ethical and the strategic ground to be defined by others.

What I Am Recommending To Defence And Aerospace Boards And Policymakers

- Adopt the House of Lords committee's recommendation for a specific, legally defined standard of meaningful human control, rather than leaving 'context-appropriate involvement' open to interpretation across different parts of government and industry.
- Treat the escalation research findings as a design constraint, not a talking point. Any system with reduced human oversight in a high-stakes decision should be assumed to default toward the aggressive option unless specifically designed and tested otherwise.
- Build assurance, testing and audit capability at the same funding priority as capability itself, rather than treating the current £4 billion autonomous systems investment as separate from the governance structure needed to oversee what it produces.

Conclusion

Every other paper in this series uses the Bengio and Hinton mechanism as an explanatory tool. In UK defence, it is the actual policy debate happening in Parliament and government right now, in 2026, with ministers on record taking visibly different positions on how much human control should remain in life-or-death decisions. Kodak took thirty seven years to find out what an unmanaged capability to governance gap costs. In this sector, the cost of finding out is not a company. It is measured differently, and it is why this paper argues the doctrine question should be resolved deliberately, now, rather than left to whichever capability or adversary forces the next public statement.

Sources: Computer Weekly, House of Lords Proceed With Caution report and UK Parliament committee evidence, RT World News, Defence Online, UK Ministry of Defence Ambitious Safe and Responsible policy statement. All figures and statements current as of mid-2026, verified at time of writing. This is a first draft prepared for internal discussion.