Chapter 2: Latency Kills – The Fatal Cost of Cloud Dependence in Combat

Part of the series: The Argument for Embedded Logic at the Edge vs Centralised Large AI in Modern and Future Warfare - April 2025

John Blamire

Founder - Ambient Stratagem

in: **linkedin.com/in/john-blamire**

The Unforgiving Pace of Battle

> "Speed is the essence of war." — Sun Tzu, The Art of War

The pace of battle is unforgiving. From urban ambushes to drone swarms, from GPS-spoofed kill boxes to dismounted troops operating under jamming, the margin between decision and destruction is often measured in milliseconds. In such a context, latency is not a technical metric—it is a battlefield risk.

This chapter presents the case that AI systems requiring remote or centralised inference introduce unacceptable delays in decision-making. These delays are not abstract—they break the kill chain, erode trust between human and machine, and can become fatal points of failure under fire.

1. How Cloud-Based AI Increases Latency

AI models operating in hyperscale cloud environments typically depend on:

- Real-time data upload from edge devices to remote servers.
- Remote inference processes carried out in distant data centres.
- Downlinked results transmitted back to the operator or platform.

In optimal conditions, this creates a round-trip latency of 100–300 milliseconds or more. In contested environments—where bandwidth is degraded or signal paths are under attack—this delay can increase unpredictably or fail altogether.

For combat operations, this delay introduces lethal risk:

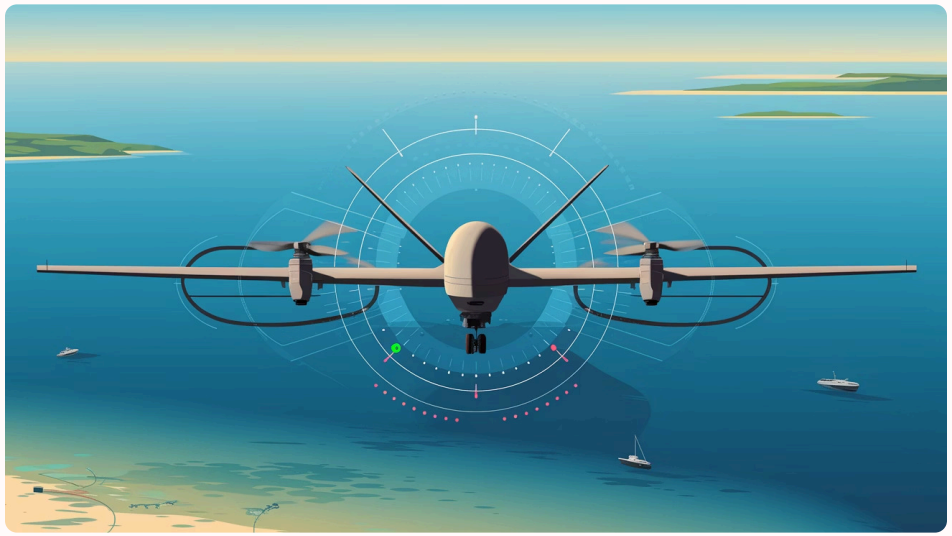
- Drone targeting logic that fails to keep up with moving targets.
- ISR feeds that lag behind battlefield movement.
- Autonomy systems that pause or loop in the absence of a server response.

2. Case Study: Red Sea, April 2025

In a joint maritime ISR mission, a US-operated MQ-9 Reaper was conducting overwatch near the Bab el-Mandeb strait. A previously reliable object recognition system—tied to a cloud-based visual model—began misclassifying small vessels due to signal interference.

Iranian-backed Houthi forces deployed portable GPS jammers and spectrum scramblers, temporarily severing the drone's uplink. In the absence of its remote inference system, the drone lost situational clarity and had to abort its objective.

This was not a failure of classification accuracy. It was a failure of architectural design—a critical AI function placed too far from the point of need.



Key Failure Points

- Cloud-dependent visual recognition
- Vulnerable uplink connection
- Lack of local processing capability
- Mission abort due to connectivity loss

3. The Tactical Cost of Delay

Let us consider the kill chain. Every additional second in a sensor-to-shooter cycle:

- Increases enemy survivability.
- Decreases the value of ISR data.
- Lowers operator trust in the AI's recommendations.

In the pressure of battle, latency:

- Breaks targeting cycles by missing fleeting windows.
- Creates ambiguity by failing to synchronise actions.
- Endangers friendly forces due to untimely or unclear decision support.

> "If you're waiting for a server farm to tell your robot what to do, you've already lost the fight."
> — US Marine Corps Robotics Engineer, 2023

4. Embedded Logic: Closing the Loop at Zero-Mile Range

By contrast, embedded logic architectures:

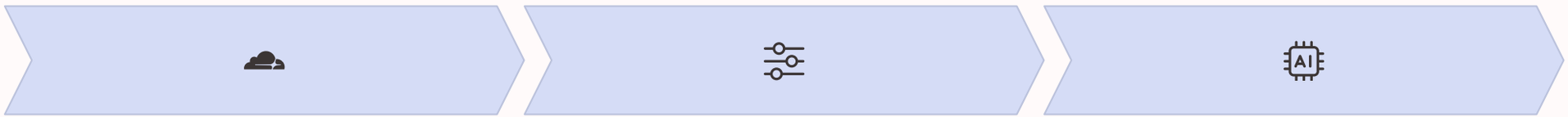
- Execute inference locally, on-chip or in-device.
- Eliminate cloud dependencies, enabling deterministic behaviour.
- Allow for tailored models that prioritise mission-specific performance over general-purpose capability.

This doesn't mean less intelligence—it means faster, more relevant, and trusted intelligence.

A drone that carries its own logic makes its own choices when cut off. A ground unit with embedded targeting support doesn't wait for permission from the cloud—it executes confidently.

This architecture enables what field commanders are now demanding:

AI that acts as a co-pilot, not a call centre.



Cloud Processing

High latency, vulnerable connections

Edge Computing

Reduced latency, improved reliability

Embedded Logic

Zero latency, autonomous operation

5. Human Operators and the Trust Gap

Cloud-based systems not only introduce latency—they fracture the human-machine relationship. Operators:

- Wait for AI support that may not arrive.
- Act on old or degraded information.
- Learn to distrust systems that "freeze" or disappear under pressure.

Embedded AI, by contrast, builds trust. It responds instantly. It degrades gracefully. It lets the human stay in control—even in chaos.

> "A commander must never hesitate because the machine is still thinking." — Field Officer, British Army Future Capability Trials, 2024

Trust Factors in Combat AI

- Response time consistency
- Degradation predictability
- Operational transparency
- Human control maintenance

Conclusion

The age of low-latency, network-pervasive warfare never materialised. The age of contested, spectrum-hostile, cyber-saturated warfare is here.

Centralised AI introduces delay. Delay disrupts mission outcomes. And delay, in combat, can kill.

The only credible response is to embed the logic where the decision must happen—at the edge, now, without hesitation.

Looking Ahead

NEXT – Chapter 3: Logic in the Line of Fire – Surviving the Spectrum War

Spectrum Warfare

Exploring how embedded systems survive in contested electromagnetic environments

Resilience

Examining defensive capabilities of edge-based AI against electronic attacks

Combat Survival

Analyzing how AI systems perform under direct fire and hostile conditions