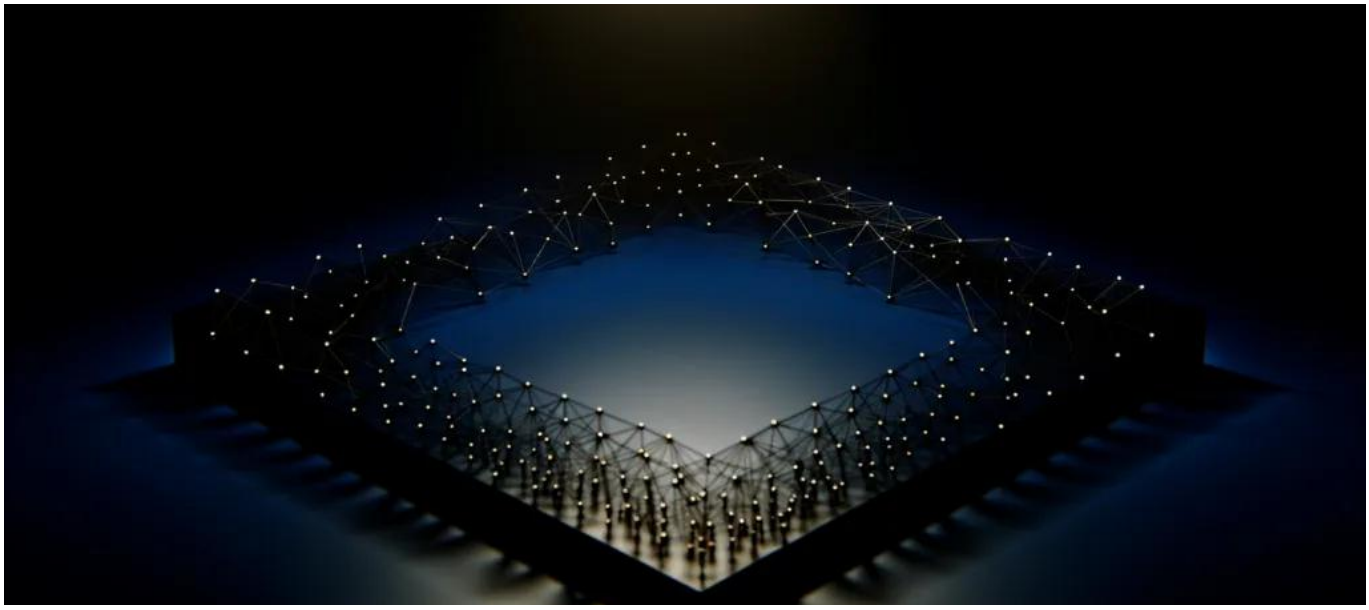

Palantir’s Maven Smart System Running on Anthropic’s Claude Powers 11,000+ US Strikes in Iran—DoD Designates It Official Programme of Record with 25,000+ Military Accounts Deployed

The Pentagon just made AI-powered killing permanent infrastructure. Palantir’s Maven Smart System, running on Anthropic’s Claude, enabled 11,000+ strikes in Iran and now holds official programme of record status with 25,000+ military accounts.

The Facts: AI Targeting Becomes Permanent Military

Infrastructure

In March 2026, the Department of Defense took an unprecedented step: it designated Palantir's Maven Smart System as an official programme of record. This bureaucratic designation carries enormous weight—it means Congressional funding is now guaranteed through September 2026 and beyond, transforming what began as an experimental project into permanent military infrastructure.

The system's combat debut came during Operation Epic Fury, which launched on February 28, 2026, with joint US-Israeli strikes on Iran. Within the first 24 hours, Maven helped identify and strike 1,000 targets. By mid-April, that number exceeded 11,000.

The numbers deserve context. In previous conflicts, target identification was a multi-day process involving human analysts reviewing satellite imagery, cross-referencing intelligence databases, and coordinating across agencies. Maven compresses this to minutes by ingesting satellite, drone, radar, and sensor data through a unified AI pipeline. The system runs on Claude, the large language model developed by Anthropic—the same AI company that positions itself as the "safety-focused" alternative to OpenAI.

More than 25,000 Maven Smart System accounts are now deployed across every US combatant command and regional theater globally. This isn't a pilot program. This is the new architecture of American military targeting.

The Technical Stack: How Maven Actually Works

Understanding Maven requires understanding the evolution from its origins. Project Maven launched in 2017 as a Pentagon initiative to apply machine learning to drone footage analysis. Google initially participated before employee protests forced the company to withdraw in 2018. The National Geospatial-Intelligence Agency formally adopted the system in 2023, and Palantir became the primary contractor.

The current Maven Smart System represents a generational leap from those early computer vision models. The architecture integrates three core capabilities:

Multi-modal sensor fusion: Maven ingests data from satellites, surveillance drones,

ground-based radar, electronic signals intelligence, and networked sensors simultaneously. The system correlates information across these streams to identify potential targets that any single source would miss.

LLM-powered reasoning: Claude serves as the reasoning layer, interpreting fused sensor data against military doctrine, rules of engagement, and historical targeting patterns. This is where the "smart" in Maven Smart System originates—the AI doesn't just detect objects, it contextualizes them within operational frameworks.

Human-on-the-loop targeting: Maven presents prioritized target packages to human operators who retain final strike authority. The interface compresses what previously required teams of analysts into decisions a single operator can make in minutes.

The architecture represents a fundamental shift in how military AI systems are designed. Previous autonomous systems like Israel's Harop loitering munitions operated on narrow, predefined parameters. Maven operates on natural language reasoning applied to complex, ambiguous scenarios.

For engineers familiar with modern ML pipelines, think of Maven as a RAG (Retrieval-Augmented Generation) system at military scale. Claude retrieves relevant tactical context from massive intelligence databases while generating targeting recommendations based on real-time sensor feeds. The system prompt likely includes operational constraints, rules of engagement, and classification hierarchies that shape every output.

The Anthropic Question: Safety-Focused AI Powers Lethal Targeting

Anthropic has built its brand on AI safety. The company was founded by former OpenAI researchers who left specifically over concerns about moving too fast on capability without adequate safety research. Anthropic's Constitutional AI methodology, its emphasis on harmlessness training, and its public communications all position Claude as the responsible choice in foundation models.

Now Claude powers a system that has enabled 11,000 strikes in seven weeks.

This isn't necessarily contradictory—Anthropic's leadership has distinguished between preventing AI from causing harm autonomously versus AI being used as a tool by legitimate authorities. But the distinction matters enormously for how the industry thinks about AI safety.

> The question isn't whether Claude can be used for military targeting. The question is whether "safety-focused AI" means anything if the safety is only about the AI's own behavior, not its downstream applications.

Anthropic reportedly implemented custom Constitutional AI constraints for military applications—guardrails designed to prevent certain categories of targeting recommendations. But the details remain classified. What we know is that Maven, running on Claude, helped identify over 11,000 targets, including the Iranian girls' school strike that killed over 170 children and is now under Pentagon investigation.

For technical leaders evaluating AI partnerships, this raises uncomfortable questions. Enterprise AI contracts typically include acceptable use policies. Anthropic's policies permit government and defense applications under certain conditions. But when your model helps select targets that result in mass civilian casualties—regardless of whether the fault lies with the AI, the operators, or the intelligence inputs—the "safety-focused" branding becomes complicated.

The tweetable insight: Anthropic built the "safety-first" AI company. Its model now powers the most lethal AI targeting system ever deployed. Both things are true simultaneously.

What Most Coverage Gets Wrong

Media coverage of Maven falls into two predictable camps: breathless warnings about autonomous killer robots, or triumphant coverage of American technological superiority. Both miss the actual story.

Wrong take #1: Maven represents autonomous weapons. Maven is explicitly human-on-the-loop. Every strike requires human authorization. The AI recommends; humans decide. This matters legally under international humanitarian law and practically for understanding

the system's failure modes. When 170 children die in a school strike, the question isn't whether an AI went rogue—it's whether the AI's recommendations were accurate, whether humans questioned those recommendations, and whether the compressed decision timelines Maven enables leave adequate room for deliberation.

Wrong take #2: This is unprecedented AI warfare. AI has been embedded in military systems for decades. Patriot missile batteries use AI for threat assessment. AEGIS combat systems incorporate machine learning for target tracking. What's new about Maven isn't AI in warfare—it's LLM reasoning applied to targeting decisions at operational scale. The shift from narrow computer vision to general-purpose reasoning models represents a qualitative change in capability, but framing it as "AI warfare begins" misses the long evolution that preceded this moment.

Wrong take #3: The civilian casualty problem is an AI problem. The Iranian girls' school strike is under investigation, but early reporting suggests the intelligence inputs identified the building as a military command facility. If that's accurate, the failure isn't Maven hallucinating a target—it's upstream intelligence being wrong. AI systems inherit the biases and errors of their training data and inputs. A perfect targeting AI given bad intelligence produces bad targeting.

The underhyped story is the decision-time compression itself. When target identification takes days, there's time for verification, questioning, and deliberation. When it takes minutes, the institutional pressure to act on AI recommendations intensifies. Human-on-the-loop becomes human-rubber-stamping-the-loop when operators face queues of AI-generated targets and implicit expectations to process them rapidly.

The Program of Record Implications

The "programme of record" designation deserves more attention than it's received. In Pentagon acquisition terms, this status means:

- **Guaranteed multi-year funding:** Maven is no longer dependent on annual discretionary allocations or demonstration project renewals. Congressional appropriations flow automatically.
- **Institutional permanence:** Killing a programme of record requires active political

effort. Maintaining it requires nothing. Maven is now default military infrastructure.
- **Contractor lock-in:** Palantir's position as primary contractor becomes increasingly difficult to challenge. The switching costs for 25,000+ accounts across every combatant command are prohibitive.
- **Expansion authority:** Programme of record status typically enables scope expansion without requiring new acquisition processes. Maven can grow capability without returning to Congress.

For defense contractors watching this space, the Maven designation signals that AI targeting systems have crossed from experimental to foundational. The Air Force's new operational AI wargame system suggests the services are building entire training pipelines around AI-assisted decision-making.

For AI companies considering defense contracts, the calculus has shifted. The market isn't speculative anymore. The DoD has demonstrated willingness to move AI systems from prototype to permanent infrastructure within a few years. The question becomes whether your company wants to compete for that business—and whether you can navigate the ethical, reputational, and technical challenges it brings.

Technical Implications for Enterprise AI Architectures

Maven's architecture offers lessons for non-military AI deployments, particularly for organizations building decision-support systems at scale.

Multi-modal fusion is the future

Maven's ability to correlate satellite imagery, radar signatures, signals intelligence, and sensor feeds into unified assessments represents where enterprise AI is heading. Financial services firms need similar capabilities for fraud detection across transaction patterns, communication metadata, and behavioral signals. Healthcare systems need multi-modal fusion across imaging, lab results, clinical notes, and genomic data.

The technical challenge isn't any single modality—it's the fusion layer. Maven reportedly uses Claude for this reasoning synthesis, suggesting that general-purpose LLMs can serve as "connective tissue" across specialized models. This architecture pattern—specialized

models for each modality feeding into an LLM reasoning layer—may prove more practical than training massive multi-modal models from scratch.

Human-in-the-loop design requires time budgets

Maven's compression of targeting from days to minutes creates a new problem: human oversight that's technically present but practically absent. Any enterprise deploying AI decision-support must explicitly design for human deliberation time.

If your AI system generates recommendations faster than humans can meaningfully evaluate them, you haven't built human-in-the-loop oversight. You've built an AI approval factory with human staging. The Maven lesson is that technical capability to keep humans in the loop differs from organizational capacity to exercise meaningful oversight.

Model selection has downstream implications

Palantir chose Claude for Maven. That choice carries consequences beyond model capability—it shapes what the system can and cannot do, what constraints apply, and what accountability exists when things go wrong.

Organizations building high-stakes AI systems must evaluate foundation model providers not just on benchmarks but on governance. What are the provider's policies on downstream use? What visibility do they have into how their models are deployed? What happens when a deployment causes harm?

The Geopolitical Compute Race

Maven's operational success will accelerate global AI weapons development. China has been pursuing similar capabilities and now has proof of concept that LLM-powered targeting works at scale.

The implications extend beyond weapons. Maven demonstrates that whoever controls advanced AI infrastructure controls the tempo of modern conflict. Target identification in minutes versus days isn't just faster—it enables different kinds of operations entirely. The 1,000 targets in 24 hours during Epic Fury's opening would have been logistically

impossible without AI-powered identification.

This creates strategic pressure on AI development that transcends commercial competition. Nations that fall behind in AI capability face potential adversaries who can identify, process, and strike targets faster than they can respond. The result is an AI arms race that may prove harder to constrain than nuclear weapons because the dual-use nature of AI technology makes verification nearly impossible.

For technology leaders, this means AI infrastructure increasingly carries national security weight. The chips that train these models, the cloud infrastructure that runs them, and the talent that builds them are becoming strategic assets subject to export controls, foreign investment restrictions, and government interest of a kind that didn't apply to software before.

Practical Guidance for Technical Leaders

Given Maven's emergence as permanent military infrastructure, what should CTOs, senior engineers, and tech founders actually do?

Audit your AI supply chain

If you're using Claude, GPT-4, or any major foundation model, you're using infrastructure that may also power military applications. This isn't inherently problematic, but it requires conscious risk assessment. Your customers and employees may have views on this. Your investors may have questions. Know your exposure and have answers ready.

Evaluate defense market opportunities soberly

The DoD has demonstrated appetite for AI systems and willingness to move quickly from prototype to programme of record. If you're considering defense work, understand that the market is real but comes with genuine ethical weight. The school strike investigation will have consequences for contractor liability and accountability. Defense AI isn't normal enterprise software.

Design human oversight with actual time budgets

If you're building decision-support AI, the Maven experience shows that technical human-in-the-loop isn't the same as meaningful oversight. Build explicit deliberation time into your system design. Measure not just whether humans approve AI recommendations, but whether they have sufficient information and time to override them.

Plan for regulatory change

The 170 children killed in the school strike will generate policy responses. The EU AI Act already imposes requirements on high-risk AI systems. US regulation is coming, particularly for AI in consequential decisions. Design your systems with auditability, explainability, and accountability in mind. The organizations that build these capabilities proactively will adapt more easily than those forced to retrofit them.

Watch Palantir's commercial moves

Palantir has historically leveraged government contracts into commercial offerings. The company now has battle-tested AI decision-support infrastructure at 25,000+ user scale. Expect commercial versions targeting enterprise decision-making in supply chain, financial services, and healthcare. Whether you view Palantir as partner, competitor, or cautionary tale, their capabilities deserve attention.

Where This Goes: 6-12 Month Outlook

Several developments are predictable from Maven's programme of record status:

Congressional scrutiny intensifies through summer 2026: The school strike investigation will produce hearings. Expect testimony from Pentagon officials, Palantir executives, and potentially Anthropic leadership. The permanent funding status taking effect in September makes the timing politically loaded.

Allied nations acquire similar capabilities by early 2027: The UK, Australia, Japan, and NATO allies will seek access to Maven or equivalent systems. This creates export control questions and potential for a broader AI targeting ecosystem across aligned nations.

Competing systems emerge from other defense primes: Lockheed Martin, Raytheon, and Northrop Grumman will accelerate their own AI targeting development. The programme of record status validates the market; competitors will follow.

AI model providers face harder policy questions: Anthropic's position as the foundation model behind 11,000 strikes will force every major AI lab to clarify its defense policies. OpenAI's recent defense contract moves suggest the industry is moving toward engagement rather than abstention. Google's 2018 Project Maven withdrawal looks increasingly like historical exception rather than template.

International law struggles to adapt: The existing framework for autonomous weapons assumes systems that target independently. Maven's human-on-the-loop design doesn't fit cleanly into existing categories. Expect diplomatic efforts to address AI-assisted targeting that blur without resolving the underlying questions.

The Fundamental Trade-off

Maven forces a conversation the tech industry has avoided. AI systems that reason generally across complex situations can be used to do genuinely beneficial things—medical diagnosis, scientific research, accessibility tools—and to identify 11,000 targets in seven weeks. The same capability serves both uses.

The industry's implicit assumption has been that we can develop powerful reasoning systems while controlling their applications through policy, terms of service, and responsible deployment practices. Maven tests that assumption. Anthropic's safety research didn't prevent Claude from powering military targeting; it may have shaped how that targeting works, but the fundamental capability remains available for this use.

There's no clean answer here. Refusing defense contracts means ceding the space to competitors with potentially less emphasis on safety. Engaging with defense means accepting that your AI will be used for lethal purposes, regardless of whatever guardrails you implement. The middle path—engaging while maintaining meaningful constraints—requires trust in institutions that have histories of exceeding their authorities.

For technical leaders, the practical implication is that AI ethics can no longer be abstract. If

you build powerful AI systems, those systems will be used in ways you don't choose and may not prefer. Your responsibility extends to thinking through those uses, designing accountability mechanisms, and accepting that your technical work has moral weight.

Maven is an official programme of record now. The AI-powered targeting capability is permanent infrastructure. The debate about whether this should happen is over; the debate about what comes next is just beginning.

AI targeting has moved from experimental to institutional—the companies that build, deploy, and govern these systems now carry responsibility for how warfare evolves.

WRITTEN BY

Artur Markus

Software engineer in Basel building AI infrastructure and agentic systems, with a background in pharmaceutical automation, GxP compliance, and validation. I help teams turn AI from a chatbot into a reliable operational layer.

Book a meeting