Defense Alpha — Project Overview

February 2026

What It Is

Defense Alpha is an intelligence engine that identifies which small defense technology companies are most likely to win contracts, raise capital, or get acquired — before it's obvious.

It does this by aggregating three public but disconnected government data sources, resolving them into a unified entity graph, and running signal detection and policy alignment scoring across the full universe.

The Data Foundation

We ingest and cross-reference three federal datasets that, individually, are public but messy and disconnected:

- **DoD Contracts** (USASpending) — 13,340 contracts worth $1.16T. Who's winning money from which agencies, at what velocity.
- **SBIR/STTR Awards** — 29,500+ R&D awards to small businesses. The earliest indicator that a company has real technology and government traction.
- **SEC Form D Filings** — Private capital raises. Shows which defense startups are attracting VC money and when.

These get deduplicated and linked through entity resolution — matching companies across sources using CAGE codes, DUNS numbers, EINs, and fuzzy name matching. The result is a unified view of ~10,200 entities: who they are, what they build, who funds them, and how fast they're growing.

What Makes It Smart

Business Classification. Every SBIR entity is classified by what they *build*, not what technology they *touch*. A company that builds door locks with mesh radios is classified as "other" — not RF hardware. This sounds simple but it eliminates the false positive problem that plagues keyword-based defense market research. We classify 1,038 SBIR-funded startups across seven categories: software (42%), components (31%), aerospace platforms (9%), RF hardware (5%), services, systems integrators, and other.

Signal Detection. 14,075 active signals across 13 types. These include SBIR phase transitions, SBIR-to-VC funding sequences, multi-agency interest, contract growth velocity, and risk indicators like customer concentration and stalled SBIR programs. Each entity gets a composite score that balances positive momentum

against risk flags.

Policy Alignment Scoring. Every entity is scored against 10 national defense priority areas — space resilience, autonomous systems, electronic warfare, JADC2, etc. — weighted by actual FY2026 budget growth rates. Space is +38%, autonomous systems +22%, hypersonics is -43%. A company building radar for space gets a policy tailwind. A company in hypersonics gets a headwind. Nobody else is doing this systematically.

Outcome Tracking. We've started recording which flagged companies actually win new contracts, raise funding, or go quiet — the beginning of a backtesting dataset that measures whether our signals predict real outcomes. This data gets more valuable every week and cannot be replicated by someone who starts later.

What It Produces

The platform generates ranked prospect reports for specific technology verticals. The first deliverable — an RF & Communications report — identified 56 verified RF hardware companies out of 1,038 classified startups, ranked by a composite of signal strength, policy alignment, and execution evidence (actual contract conversions).

The same pipeline can produce equivalent reports for any vertical: defense software, autonomous systems, space technology, electronic warfare, directed energy, etc. Each report takes hours instead of the weeks it would take a human analyst, and the methodology is transparent and reproducible.

Where the Moat Is

What's not defensible: The raw data is public. The LLM classification costs $55 to run across the entire universe. The signal detection logic is rule-based. Any competent developer could rebuild the data pipeline.

What is defensible:

- *Entity resolution at scale* — 10,200 deduplicated entities linked across three federal data sources. Tedious, error-prone work that nobody wants to redo.
- *Policy alignment calibration* — requires domain expertise to maintain the priority taxonomy and budget weights each budget cycle. Not an engineering problem — a knowledge problem.
- *Outcome tracking over time* — the longer we record signal-to-outcome data, the more accurately we can say which signals predict success. This is a time-locked asset. Starting six months from now means six months less data.
- *Workflow integration* — when users build watchlists, tag companies, and add proprietary notes, the platform becomes harder to leave. Multi-user data creates network effects.

Current State and Next Steps

The core infrastructure is built and the first vertical report has been delivered to a defense sales consultant who called the results "all new companies to me." The immediate roadmap:

1. Deliver two more vertical reports (defense software, space technology) to targeted contacts in VC and defense industry
2. Build the funding raise outcome detector and begin weekly automated data refreshes
3. Backtest signal accuracy after 3-6 months of outcome data accumulates
4. If multiple users ask for self-serve access, build the dashboard

The strategic positioning is "defense intelligence for private markets" — the same analytical edge that Palantir sells to governments, applied to the investment and BD layer of the defense ecosystem.