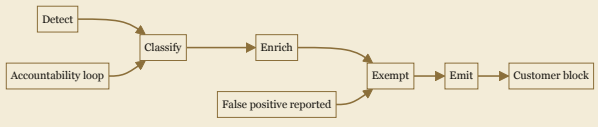
Methodology · The Short Version · XI April MMXXVI

# How We Built a Threat Feed That's Faster and More Accurate Than the *Billion-Dollar Vendors*

*The short version. Three minutes. Five principles. Two math tricks. One NDA-gated full methodology at the bottom.*

> **Today we ship threat intelligence to 275+ organizations in 46 countries, running on about $500 a month of Azure compute, with an internal site-level false-positive rate under 0.004 per cent.** *CrowdStrike's cheapest Falcon Intelligence tier is around $100,000 per year. Recorded Future's enterprise plan is $50,000+ per seat. Mandiant Advantage starts at $75,000. We charge zero for the free tier and $45 per month for Starter. Our ingestion-to-publication latency is under ten minutes. Theirs is 24 to 72 hours. The difference is not scale. It is architecture.*

This is the short version. The full methodology is an 8,500-word whitepaper with flow diagrams for every detection, classification, enrichment, exemption, and alert path in the platform. We share it under NDA with customers, investors, and security teams evaluating us for partnership. The email is at the bottom of this document. What follows is the compressed version.




Five stages, each running on its own cadence, each correctable in place, each with its own measurable accuracy number. Most commercial feeds conflate two or three of these stages into a

single monolithic pipeline, which is why correcting a false positive in a commercial feed takes weeks. When detection, classification, enrichment, exemption, and emission are separate layers, a bad signal can be corrected at any one of them without tearing down the others.

## Five Principles That Do the Actual Work

I. **Separate the layers.** Detection, classification, enrichment, exemption, and emission are five distinct stages. Each has its own inputs, its own outputs, its own cadence, and its own error budget. When you conflate them, every correction becomes a rebuild. When you separate them, every correction is a config change.

II. **Parallelize everything.** Every ingestion source, every classifier, every enrichment call, every AI model query, every cron job runs concurrently. Our 15-vendor AIPM audit published this morning completed in 33.2 seconds end-to-end because five models queried four dimensions across fifteen domains in parallel. Sequential code in a threat pipeline is a performance bug.

III. **Use AI only where AI is best.** LLMs are good at language, ambiguity, and research. They are bad at deterministic scoring, edge-case rules, and fast lookups. A pipeline that uses an LLM for every decision is guaranteed to be both slower and less accurate than one that uses LLMs only at the decision points where nothing else works. Discipline about where you do *not* use AI is as important as cleverness about where you do.

IV. **The exemption layer is where the accuracy lives.** Detection and classification always produce false positives. The difference between a feed with a 5 per cent false positive rate and a feed with a sub-0.01 per cent false positive rate is the quality and maintenance of the exemption layer. We put as much engineering into the "definitely not a threat" path as we put into the "definitely a threat" path.

V. **Write down every mistake.** Every confirmed false positive, every missed detection, every deploy failure, every customer complaint becomes a structured incident record, an automated compliance pattern, and a standing lesson-learned that loads into every future engineering session. Accountability is not a marketing virtue; it is an architectural feature that compounds over time.

## Two Math Tricks Nobody Else in the Category Is Doing

### I. Bloom Filter Novelty Check

A Bloom filter is a probabilistic data structure that answers one question fast: *have I ever seen this thing before?* It costs a handful of bytes per million items and returns the answer in a microsecond. We keep a continuously-updated Bloom filter over the full indicator space — over a million indicators as of this writing — and every new candidate indicator gets a novelty check *before* classification. Known-returning-bad routes one way. First-ever-seen routes another way. This is what prevents the single biggest source of false positives in any threat feed: re-scoring known-bad indicators as known-good after a tenant reassignment. The check runs in O(1) regardless of corpus size.

### II. Cross-Index Correlation in One Query

The Butterbot platform uses a single Meilisearch substrate across 42 separate indexes — IOCs, block events, threat-intel pulses, offshore entity relationships, behavioral sessions, adversary profiles, AIPM audits, and more. When a single indicator (an IP, a domain, a company name) appears in multiple otherwise-unrelated indexes within a short time window, *the correlation is itself a signal*. It catches attack campaigns in their earliest phase because the attacker's infrastructure shows up across multiple data surfaces before it shows up as a confirmed threat on any single one. Most commercial threat platforms store their data in separate databases per product line and physically cannot do this in real time. We correlate in milliseconds because we built the whole platform on one substrate on purpose.

## The Receipts, for the Skeptical

> *This week we published the first quarterly "State of AI Brand Perception in Cybersecurity" report. Fifteen named cybersecurity vendors. Five AI models. Four verbatim fabrications we caught in 33 seconds — OpenAI insisting CrowdStrike is headquartered in Sunnyvale (it's been Austin since 2022), Gemini inventing a Rapid7 founder named "Alan Chhabra," Gemini mutating Snyk's Danny Grander into an unrelated security researcher named Danny Gruss, DeepSeek confusing Wiz's Roy Reznik with monday.com's Roy Mann. Every named error is reproducible. Every audit is re-runnable. The full report and its PDF are at* ***aipmsec.com****. That quarterly report exists because the architecture this page describes makes it cheap to generate — 15 vendors, 5 models, 75 audits, 33 seconds flat, on a weekend.*

### The Full Methodology, Under NDA

The full whitepaper covers nine parallel ingestion sources, eight independent classifiers, eight enrichment cross-references, ten false-positive prevention mechanisms, the AI integration philosophy in detail, the accountability loop, a comparison table against the three biggest commercial vendors, and flowcharts for each.

If you want it, email the address below with the subject line **methodology** and one sentence about who you are. We read every one and respond within 24 hours. The NDA is one page and your legal team will not hate it.

*patrick@dugganusa.com*

---

DugganUSA LLC — Minneapolis Minnesota — Founded 2025–10–07
Short version of the internal threat evaluation whitepaper v1.0
Full document ~8,500 words · 12 sections · 7 flow diagrams · NDA-gated