Matt Coburn

AI startup founder and builder

I design and implement **document → ontology → knowledge-graph systems** with strict grounding, provenance, and evaluation invariants, and take them from design to revenue as a **hands-on player-coach** (70/30 coding/management).

Track record: Built multiple AI document platforms for Fortune 100 clients. Shipped production ML to 2 of top 3 global automakers. Generated $2M+ revenue as technical co-founder. Led teams of 15+ engineers and Data Scientists.

[Aristotle](#) — Founding Engineer & Tech Lead (80% hands-on / 20% management)

Los Angeles, CA | June 2024 – Present

Own technical direction end-to-end and shipped a zero-to-one enterprise document intelligence platform into production for customers including **2 of the top 3 global automakers**. Built the core backend, set engineering standards, and **hired/led 4 engineers** while staying deeply hands-on.

Core invariant: *no model output can change system state—or reach a user—unless it is schema-valid, auditable, and traceable to source documents.*

What I built:

Key contributions:

- **Ontology & entity system:** Deterministic core entities (People, Orgs, Events, Relationships) in Pydantic + Postgres with strict identity semantics; extensible customer-defined schemas validated at ingest
- **Document → fact → provenance pipeline:** Structured extraction with click-to-source spans/pages; ambiguity forced into explicit review flows
- **Grounded LLM UX:** Streaming answers with verified vs pending states, citations for assertions/quotes, and HITL review tools
- **Agent runtime:** Tool-using agents implemented as explicit state machines with boundary validation and retry/escalation semantics
- **Retrieval + eval gates:** Hybrid BM25 + dense retrieval tuned on Precision@K / Recall@K; regression harnesses used as release gates
- **Engineering leadership ("golden path"):** Established CI/CD, code review standards, release gates, observability, and production-readiness rituals. Personally built foundational scaffolding (auth, service patterns, shared model contracts), then used it to unblock and scale delivery cadence.

Impact: Enabled enterprise workflows that reduced manual document review, enforced grounding guarantees, and unblocked adoption in regulated environments.

[WorkFusion](#) — VP of Data Science (Applied AI & Document Intelligence)

New York, NY | 2023

Joined as a **hands-on VP** to ship applied AI in regulated financial environments, leading **~15 data scientists** while staying hands-on in system design and core implementation.

- Shipped **KYC/AML systems** (transaction monitoring, risk scoring, case prioritization) for major institutions; improved detection outcomes by **~35%**.
- Built **LLM-based document extraction & auto-labeling** for high-volume IDP workflows with strict governance
- Designed security-first MLOps, evaluation, and audit practices for regulated production

Tangible Intelligence — Founder / Chief Data Scientist / CEO

Dallas, TX | Jan 2020 – Apr 2023 Technical co-founder and CEO of a document-intelligence startup; generated **~$2M enterprise revenue** over two years.

- Built a **no-code HITL extraction platform** with ontology-driven schemas, deterministic entity resolution, and click-to-source provenance
- Built **SafeScan**, a real-time sensitive-data and anomaly detection system combining ML, rules, and search
- Led SOC 2–aligned deployments; scaled and managed an **8-person** engineering/data team

M Science (Jefferies subsidiary) — Principal Data Scientist, Founding Lead (Ontology Platform)

New York, NY | May 2018 – Dec 2019

Conceived, architected, and shipped an ontology-driven knowledge platform as an internal startup. Hired and led an independent team, shipped to production, and supported clients including **BlackRock, Two Sigma, and Citadel**.

- Built an **Ontology-as-a-Service** layer modeling companies, securities, events, and relationships across public + proprietary data.
- Designed **entity resolution** rules and relationship semantics prioritizing deterministic identity and reproducibility over heuristic joins.
- Built an application layer that let clients apply shared ontology semantics to private datasets with **isolation and fine-grained access control**.

Expedia Group (Hotels.com) — Data Scientist (Search & Ads Optimization)

Dallas, TX | Oct 2016 – May 2018

Worked on large-scale search/ads systems with models and pipelines influencing **$200M+ annual spend**.

- Built models estimating expected value of clicks across **millions of keywords**, enabling daily automated bid optimization.
- Identified systemic underperformance across publisher inventory; reallocated spend, yielding **~240% ROI improvement** on targeted segments.
- Built revenue-critical production pipelines with monitoring and safeguards.

Technology Stack

Core: Python, SQL, C, TypeScript, Zig
Backend: FastAPI, Pydantic, PostgreSQL, Redis, Docker, Kubernetes, AWS
Data & Infra: Pandas, NumPy, Spark, ETL/streaming pipelines, vector DBs
LLMs & NLP: PyTorch, HuggingFace, OpenAI/Anthropic APIs, RAG architectures, hybrid retrieval (BM25 + embeddings)
Specialties: Document intelligence, structured extraction, ontology design, AIUX, evaluation frameworks, production ML governance

Education

University of Texas at Dallas

- B.S. - Electrical Engineering, 2013
- Masters Coursework - Computer Science, 2016

Let's Build Something Extraordinary

I'm passionate about making AI work reliably in production with deterministic guarantees. Looking for Staff/Principal Engineering or Applied AI Leadership roles where I can leverage my experience building enterprise-grade ML systems.