Thomas Ferrell

https://taf.codes

Summary

Generalist across infrastructure, backend, and beyond. 12 years of experience at mostly startups, twice as a founding engineer. Comfortable across most abstraction layers, my capabilities extend down into firmware and up to a little bit of frontend. Broad background, and eager to learn new domain-specific knowledge. Recent focus on applied AI: embeddings, semantic search, evaluation, and local inference. Self-directed by default, and at my best with a hard problem and room to run.

Skills

Languages & Frameworks: Python, Elixir/Phoenix, JavaScript, Flask, Ruby, Java, SQL, HTML/CSS, Shell/Bash, C

Infrastructure: AWS, Docker, GNU/Linux, Serverless, DigitalOcean, Infrastructure-as-Code

Data & AI: LLM Orchestration & Evaluation, Embeddings, Vector Search, Local GPU Inference, Whisper, vLLM

Tools: Git, PostgreSQL, pgvector, DuckDB, NoSQL, Playwright, Promptfoo, Label Studio, Wireshark, Jupyter, etc

Ready to expand into: Kubernetes, GCP, Rust

Work Experience

FetchFox AI

Founding Software Engineer — 2025

- Architected scalable infrastructure and backend powering realtime, user-driven web data extraction
- Designed a tiered routing system for network and browser resources to improve both cost-efficiency and latency
- Authored, documented, and published the FetchFox SDK Python package

Out&Back Outdoor

Senior Software Engineer — 2022 - 2024

- Medium-scale web extraction of product data; built adaptable pipelines to clean and ingest messy third-party data (ETL)
- Shipped features end-to-end across multiple large projects (> 1 million LOC)

Proxor Inc

Founding Engineer — 2016 - 2021

- Architected and shipped a cloud exam-delivery platform on AWS end-to-end: browser-based virtual desktops, serverless real-time booking and provisioning, and integrated third-party scheduling and proctoring
- Owned global multi-region deployment via infrastructure-as-code, serving thousands of examinees across many countries
- Built automated grading for Java, JavaScript, Python, and C; evolved it into a scalable production service
- Unified an on-prem LAN-based exam delivery with the cloud offering; partnered with IEEE on a production launch

Independent Technical Consulting — 2013 - 2016

- Built custom Linux exam environments and DevOps tooling; collaborated with CS faculty at MIT and CMU

Carnegie Mellon University, School of Computer Science — Technical Assistant, 2012 - 2013

Selected Projects & Writing

detailed writeups at https://taf.codes

Podcast Trend Analysis

Elixir/Phoenix app surfacing trends across hundreds of millions of words of transcribed podcast audio, scoring relevance over embedding space by feed and time. Postgres + pgvector with search; currently migrating to a columnar store to better suit OLAP queries.

LLM Evaluation & Classification Pipeline

LLM classifier that removes advertisements from podcast content, cleaning the data supporting the trends above. Scored several local models and a frontier model on a hand-labeled set by precision/recall/F1; a small local model matched it at ~1/100th the cost.

Personal Knowledgebase

Automated intake pipeline that fetches, renders, and extracts clean text from articles I email myself. Pluggable design makes it trivial to try new chunking strategies and embedding models, each computed and benchmarked automatically.

Foot-Pedal USB Speech-to-Text Peripheral

USB-HID device (RP2350, C / Pico-SDK) that streams microphone audio to a speech-to-text model and types the result via emulated keyboard, triggered by a foot-switch for hands-free use.

Education & Activities

University of Pittsburgh (2013-2015): Founding Officer, Computer Science Club

Miscellaneous: Eagle Scout, CMU Computer Club, Cycling, Metal Fabrication