

Josh Zirena

AI Tech Lead | AI Platforms & Backend Engineering

Miami, Florida · zirena.ai

SUMMARY

AI Tech Lead with 9 years of engineering experience, focused on enterprise AI platforms, large language model (LLM) evaluation, and distributed backend systems. Led teams of up to 6 engineers and delivered work across team boundaries in healthcare and financial services, combining hands-on development with technical direction and mentoring.

EXPERIENCE

Eli Lilly, AI Tech Lead (contract via Coforge)

September 2025 to Present

- Developed Python and Go functionality in the application service behind Lilly's Kubernetes-based AI platform, supporting Model Context Protocol (MCP) server registration, configuration, authentication, and tool routing. Serve as the technical point of contact for teams integrating with the service.
- Configure and support the Bifrost LLM gateway for model access and MCP tool routing, including failover, virtual keys, per-team budgets, and rate limits.
- Own and maintain the application's AWS infrastructure as code in Terraform.

ICF International, Tech Lead

July 2022 to September 2025

- Led 6 engineers delivering Node.js, TypeScript, PostgreSQL, and AWS services for the Centers for Medicare & Medicaid Services (CMS). Provided technical direction, code review, delegation, and mentoring, and owned the team's AWS infrastructure in Terraform.
- Replaced manual spot checks of HIV.gov's retrieval-augmented generation (RAG) assistant with a daily evaluation pipeline. Used a reference dataset and LLM-as-judge scoring, with separate retrieval evaluation and production-trace assessment, to track quality over time. Built with Node.js, AWS Lambda, Langfuse, and Weaviate.
- Cut turnaround for a set of iQIES reports from about 2 hours to under 1 minute by replacing a cron cycle with event-driven AWS SQS processing. Implemented the publishing integration in an upstream team's codebase with their review and approval.
- Added deployment safeguards that reject failing database migrations and prevent promotion until migrations have run successfully in a lower environment.

LL Flooring, Software Engineer

March 2022 to July 2022

- Improved e-commerce functionality and resolved performance issues in Salesforce Commerce Cloud, refactoring legacy code using JavaScript, jQuery, and Node.js.

IPD Analytics, Software Engineer

August 2020 to March 2022

- Designed and implemented internal and subscriber-facing tools for medical patent and healthcare analytics.
- Built data-driven platforms for insurers and hospitals using Node.js, React, GraphQL, and Apollo.

Finleap Connect, Software Engineer

September 2018 to August 2020

- Built secure open-banking APIs connecting third-party applications to bank accounts across Europe using Node.js and TypeScript microservices.

- Reduced error rates from 40% to 5% by improving Puppeteer web-scraping scripts and the API requests they fulfilled.
- Led 3 engineers based in Slovakia while contributing to distributed financial services systems.

YOC, Technical Support Manager

June 2017 to August 2018

- Engineered JavaScript integrations for advertising clients, helping publishers implement Google advertising tags and adding custom JavaScript functionality.

TECHNICAL SKILLS

- **AI platforms and evaluation:** LLMs, MCP, RAG, Bifrost, LLM-as-judge evaluation, golden datasets, Langfuse, Weaviate.
- **Languages and backend:** TypeScript, JavaScript, Python, Go, Node.js, GraphQL, Apollo.
- **Cloud and infrastructure:** Amazon Web Services (AWS), Lambda, SQS, Terraform, Kubernetes.
- **Data and frontend:** PostgreSQL, React.
- **Architecture and reliability:** distributed services, event-driven systems, observability, gateway access controls, deployment safeguards.

PROJECTS

Nestmind | Personal Project

- Built a natural-language feeding tracker that answers questions with averages and charts. Used the model to interpret requests and the database to calculate results.

EDUCATION

Florida International University (FIU)

Bachelor's Degree in Information Technology

Minor in Business Entrepreneurship