

STEPHEN PETTUS

Senior Software Engineer · Applied AI & Platform Engineering
Philadelphia, PA · stephennpettus@gmail.com · [linkedin.com/in/stephen-pettus](https://www.linkedin.com/in/stephen-pettus)

SUMMARY

Senior software engineer and former startup CTO building human-centered AI systems on production-grade distributed platforms — LLMs, retrieval-augmented generation, embeddings, and vector search — and an NSF PhD fellow in Human-AI Interaction. A decade shipping production software across enterprise platforms and startups, now specialized in applied LLM/RAG systems. Targeting staff / applied-AI and AI-platform roles.

CORE SKILLS

AI / ML: LLMs, Retrieval-Augmented Generation (RAG), embeddings & vector search (FAISS), AI evaluation, AWS Transcribe, applied ML for incident/failure analysis

Platform & Cloud: AWS (Lambda, SQS, SAM), Azure, Docker, CI/CD (Jenkins), observability (New Relic), systems architecture, distributed microservices

Languages: Java, Python, Go, TypeScript / JavaScript, Scala, C, PHP

Frameworks & Data: Spring, React, Angular, Django, Laravel, Node.js · PostgreSQL, MySQL, Cassandra, MongoDB, NoSQL

PROFESSIONAL EXPERIENCE

McGraw Hill

Senior Software Engineer

Philadelphia, PA

Jan 2022 – Present

Building applied-AI and platform capabilities on large-scale, distributed learning systems used by teachers and students worldwide.

- Architected and shipped an AI-driven incident-analysis system that mines historical incident data across 20+ microservices to model failure patterns, auto-generate diagnostic reports, and surface actionable insights — cutting incident triage, tracking, and resolution time by 50%+ and preventing recurring outages.
- Partnered with Data Platform Group leadership to design and evaluate architectural and process improvements across distributed microservices powering remote-learning platforms serving 25M+ digital users worldwide.
- Led a cross-functional root-cause investigation of systemic content-licensing failures; architected and deployed a remediation service under tight production constraints to correct backdated licensing at scale.

Penn Mutual

Software Engineer

Horsham, PA

Jan 2019 – Mar 2022

Delivered production applications with Java/Spring and React and helped move the organization to the cloud; worked with the Big Data team on large-scale financial datasets.

- Designed and deployed Penn Mutual's first cloud-native application, establishing architectural standards later adopted across teams.
- Built production internal applications with Java Spring and React, integrating AWS services (Lambda, SQS, SAM) into scalable, production-grade systems.
- Engineered large-scale financial data pipelines with Scala and Cassandra — analysis, preprocessing, and validation — supporting downstream analytics.

Myavana LLC

Chief Technology Officer

Atlanta, GA

Jan 2015 – Dec 2019

Technical co-founder and CTO for end-to-end product development, building data-driven systems using Node.js and AWS to power AI-based hair-health recommendations.

- Designed backend data pipelines and APIs that enabled collection, processing, and interpretation of fine-grained user data, informing recommendation logic.
- Led agile planning and technical direction, balancing experimentation with production delivery of scalable, user-facing systems.

ZipCode Wilmington / Code Differently

Instructor / Teaching Assistant

Wilmington, DE

Dec 2017 – Jan 2019

Taught Java and the Spring Framework to adult learners and professional engineers, building instructional tools and helping shape CS curricula across Delaware.

- Developed hands-on labs and instructional materials grounded in real-world computing problems.
- Built internal educational tools using Java, Spring, and TypeScript to support instructional delivery and learner progress.
- Collaborated with leadership to design and adapt computer science curricula used across schools in Delaware, emphasizing accessibility and learner retention.

Creative Tech Works
Lead Software Developer

Philadelphia, PA
Dec 2014 – Jan 2016

Led full-stack development for community-centered STEM initiatives and technology-driven outreach.

- Led development of multiple full-stack applications supporting community-centered STEM initiatives.
- Co-designed and executed technology-driven outreach events (e.g., a City Streets race and a technology-based fashion show) that integrated computing into culturally relevant contexts to increase STEM engagement.

SELECTED PROJECTS

Monday — AI-Assisted Life Operating System · *In development*

- A system to help people intentionally manage their lives, goals, and responsibilities with AI, designed around human context rather than raw automation — exploring how LLMs and agents can reduce complexity in everyday life management, dogfooded as a personal operating system and connected to my research in human-centered AI.

Snippet — AI Video Learning Platform · *Independent*

- An AI system that turns educational video into a personalized, searchable learning experience — transcribing and tagging content and using LLMs and embedding-based retrieval to surface timestamped snippets and adaptive learning plans; investigates human-centered evaluation for interpretability, trust, and equitable access.

Avina — AI-Assisted Reflective Learning & Journaling · *Independent*

- Built an end-to-end input-to-insight pipeline (AWS Transcribe, retrieval-augmented generation, LLMs) grounding outputs in curated psychological and educational knowledge; embedded structured thematic knowledge with FAISS vector search. Serves as an applied evaluation environment for my doctoral research on cognitive-state-aware AI.

EDUCATION, FELLOWSHIP & RECOGNITION

Drexel University
Ph.D., Information Science — Human-AI Interaction (in progress)

Philadelphia, PA
Starting Fall 2026

Temple University
B.S., Computer Science

Philadelphia, PA
2014 – 2018

- NSF CSGrad4US PhD Fellow (2024, 3-year award) — 3-year doctoral fellowship funding a Ph.D. in Information Science at Drexel, focused on Human-AI Interaction and AI-assisted cognition.
- Technical.ly Philadelphia RealLIST Engineers, 2022 — among 20 influential technologists building, shipping, mentoring, and community organizing.

ADDITIONAL INFORMATION

Program Management: SDLC, Systems Architecture, Jira, CI/CD pipelining

AI / ML: LLMs, Retrieval-Augmented Generation, embeddings & vector search (FAISS), AI evaluation, AWS Transcribe

Engineering: Java, Python, Go, TypeScript, Scala, Spring, React, Angular, Django, Node.js, Docker, AWS, Azure, PostgreSQL, Cassandra, MongoDB