

GIUSEPPE PANETTA

Akron, Ohio | (201) 660-4200 | jlpchetta1681@gmail.com

linkedin.com/in/jlpchetta1681 | github.com/Jojobeans1981 | portfolio-2025-blond.vercel.app

Full-Stack Software Engineer | AI/LLM Specialist | 6+ Years Experience

PROFESSIONAL SUMMARY

Full-stack engineer with 6+ years in software development and 3+ years specializing in AI/LLM integration. I build production LLM pipelines — the orchestration, the validation layer, and the latency budget — not demos. Track record shipping web, backend, and AI infrastructure, with depth in multi-LLM orchestration, prompt engineering, and autonomous software factories.

PROFESSIONAL EXPERIENCE

Co-Founder & Lead Developer — FutureEng LLC

Jan 2024 – Present · Austin, TX

- Built end-to-end automated legal intake system for Crain & Wooley law firm with Next.js, TypeScript, Supabase, Stripe and Twilio — currently processing real client intake workflows
- Architected an AI training platform for NYC Transit dispatcher-exam candidates, converting 500+ pages of legacy PDFs into a working academy in a 21-day sprint; generates dynamic quizzes and scenario-based training via multi-LLM orchestration
- Implemented SWITCHBOARD, an LLM failover router across Anthropic Claude, Groq, Gemini and Ollama that optimizes cost and latency against quota constraints
- Stack: Next.js 15, TypeScript, Tailwind CSS, Supabase/PostgreSQL, Prisma ORM, Stripe, Twilio, Claude SDK, AWS Lambda

AI Education Engineer — Superbuilders AI

Feb 2026 – Present · Remote

- Built an AI curriculum generator that automates educational content creation; deployed at Alpha Schools and Texas Sports Academy serving 100+ students with personalized learning paths
- Uses prompt engineering and multi-LLM inference to generate literacy-focused curriculum tailored to grade level and learning style

AI Software Engineer Fellow — Gauntlet

Jan 2025 – Jan 2026 · Remote

- Built 50+ AI-powered applications and prototypes against Claude, Gemini, Groq and DeepSeek APIs
- Developed an autonomous software factory that generates production-ready applications from natural-language specs, with DAG execution planning, build verification and an auto-fix loop
- Worked the real trade-offs across competing providers: multi-LLM orchestration, cost optimization, and latency

Software Instructor & Curriculum Developer — Veroskills LLC

Jan 2022 – Dec 2025 · Remote

- Tutored 500+ software engineering learners across full-stack web development, React, Node.js and database design
- Completed a 6-month curriculum development internship; offered a full-time role after 2 weeks based on performance and mentorship impact
- Designed and delivered live coding sessions, code reviews and debugging workshops

Founder & Lead Architect (*Shelved 2025*) — Prometheus Workforce Analytics

2024 – 2025 · Remote

- Designed and built a workforce analytics platform with real-time data processing and predictive modeling
- Shelved to take the Gauntlet fellowship — the product-market validation work informed later projects

AI Data Annotation & Evaluation Specialist — Alignerr, Outlier AI, Labelbox

2023 – Present · Remote

- Performed data annotation and AI evaluation for major tech companies training foundation models and LLMs
- Evaluated model outputs for accuracy, safety, coherence and alignment against company guidelines
- Contributed to training datasets for leading AI systems — direct exposure to model behavior and failure modes

SELECTED PROJECTS

- Mirra** — Real-time AI video tutor github.com/Jojobeans1981/Mirra
- Four-service streaming pipeline — Deepgram STT to Groq inference to Cartesia TTS to a lip-synced Simli avatar — over WebSocket/WebRTC at ~387ms end-to-end, measured by an instrumented endpoint rather than estimated
 - The engineering was backpressure, stream teardown and keepalive handling, not prompting; TTS first byte (~580ms) missed its budget and is documented rather than hidden

TypeScript · WebRTC · WebSockets · Groq · Deepgram · Cartesia

- GRIMOIRE** — RAG over legacy Fortran github.com/Jojobeans1981/legacylens
- Makes BLAS, LAPACK and ScaLAPACK codebases queryable in natural language — decades-old numerical linear algebra with no modern documentation
 - Naive line-window chunking split a SUBROUTINE in half and returned an algorithm with no signature; chunks are cut on routine boundaries instead, with chunk size raised 400 to 1500 chars. 15-test suite covering chunker and retrieval

Python · RAG · Vector Search · LLM

- AgentForge** — Healthcare AI agent github.com/Jojobeans1981/langchain-healthcare-tools
- 9 LangChain tools with 5-layer output verification and a 100-case evaluation dataset; integrates with OpenEMR
 - Docker, CI and a full test suite — built to be run by someone other than me

Python · LangChain · Docker · pytest

TECHNICAL SKILLS

Languages & Frameworks	JavaScript, TypeScript, Python, React, Next.js, Node.js, Express
AI & Machine Learning	LLM Integration, Prompt Engineering, Multi-LLM Orchestration, RAG Systems, AI Evaluation, Fine-tuning
Backend & Database	PostgreSQL, Supabase, Prisma ORM, AWS Lambda, AWS S3, Docker
API & Integrations	Stripe, Twilio, Claude SDK, OpenAI API, Gemini API, Groq API, MCP
Frontend & Styling	Tailwind CSS, Responsive Design, Component Architecture
Testing & Tooling	Playwright, Jest, pytest, Git/GitHub, Vercel, REST APIs

EDUCATION & CERTIFICATIONS

App Academy, 2025	Full-Stack Web Development Certificate
Lambda School (BloomTech), 2021	Full-Stack Web Development Certificate