Leon Shala

Software Engineer | Applied AI

Pristina, Kosovo · shala_leon@outlook.com · +383 45 300 675 · leonshala.dev · LinkedIn

I'm a full-stack engineer with 5+ years in healthcare and e-commerce. I've led a small engineering team and now work on ClinicOS, Aios Medical's clinical platform. My main focus is building and evaluating its AI agent for patient support.

Experience

Software Engineer (Applied AI) Aios Medical

Remote · 2026–Present

- Build tools for a patient support AI agent on a team that handles thousands of messages a day. The agent checks orders, issues refunds, cancels subscriptions, and hands cases to the care team. Its daily automation rate is 92%. The team opens 1.82 new tickets per 100 patients who contact support.
- The agent completes requests and replies about 5× faster than human agents, with better reply quality and tone and higher patient NPS.
- Build search with HyDE and Pinecone so the agent can find relevant help articles and past support cases. Trace tool calls in Langfuse. Test the agent on 748 de-identified conversations to check which actions it takes and how it replies.
- Build the ClinicOS dashboard in Next.js and React and its backend services in NestJS and GraphQL. Run background jobs with Redis and BullMQ and store application data in PostgreSQL.

Technical Lead Dardania Chamber of E-Health

Pristina · 2024–2026

- Designed the caregiving platform's frontend and backend architecture as the project's first engineer.
- Set up Azure development and production environments and built CI/CD pipelines for deployments.
- Led the team that built the public website, client web app, and multilingual admin dashboard.

Software Engineer Stree LLC

Remote · 2023–2024

- Built features for PlaySharper's Vue frontend and connected them to its FastAPI backend. Fixed layout and accessibility issues across mobile and desktop.

JavaScript Engineer Gjirafa Inc.

Pristina · 2021–2024

- Led the team that built Gjirafa's first mobile app, the GjirafaMall shopping app.
- Redesigned the GjirafaMall and Gjirafa50 online stores and restructured their frontend code to improve page performance and SEO.
- Built and maintained features for the Codex CMS admin dashboard with design and backend teams.

Technical skills

AI agents	Agentic loops, tool calling, response validation, human handoff, model fallback
Models & prompts	OpenAI, Anthropic, Gemini, prompt engineering
Retrieval & evals	RAG, HyDE, Pinecone, reranking, conversation and tool-call evals, Langfuse tracing
Web & mobile	TypeScript, JavaScript, React, Next.js, Vue.js, React Native, Expo
Backend & data	Node.js, NestJS, Express, GraphQL, REST, Prisma, PostgreSQL, MongoDB, Redis, BullMQ
Testing & delivery	Jest, Playwright, Docker, Azure, CI/CD with Azure DevOps

Education

Bachelor of Science in Computer Science

2019–2023

UBT College – Higher Education Institution, Pristina, Kosovo