

Nikhilesh Yadav

📍 Bengaluru, India ✉ nikhileshofficial@gmail.com ☎ +91 96253 73185 🌐 yniks.github.io in yniks
🔗 yniks

Experience

Senior Software Engineer

PersanaAI (YCW23)

San Francisco, USA

June 2024 – Present

- Led the design and development of the flagship “**Quantum Agent**” using **Node.js** and **Golang**, processing **20M+ monthly requests** with high reliability. Built end-to-end automation for **onboarding** and **sales workflows**, integrating **30+ data sources** including **web scraping**, dynamic email triggers, and fault-tolerant orchestration using **Temporal.io**. The system leveraged **OpenAI/Claude**, **Langchain**, **AI SDK**, and the **Mastra network layer** for context-aware LLM operations.
- Delivered a production-ready **AI chat platform** for **B2B SaaS clients** to support a key funding round, enabling fast, document-grounded conversations by extracting **insights and narratives** from **structured and unstructured customer and sales data**. Designed modular **AI agents** for **insightful search**, **query refinement**, and **workflow execution**, powered by a layered **RAG** and **ReAG** architecture for **context-aware responses** and **step-by-step reasoning**.
- Migrated backend from an **Express monolith** to modular **NestJS** and **Golang services** orchestrated by **Kafka**, enabling faster iteration, cleaner deployments, and improved code ownership. Replaced **BullMQ** with **Temporal.io**, increasing workflow reliability and reducing failure rates by **40%**.
- Rebuilt the spreadsheet UI using **React** and **Chakra UI**, reducing P99 load time from **5.43s to 1.78s**. Added features like **live editing**, **drag selection**, and instant previews. Backend tuning improved throughput to **1K req/s** and cut API costs by **25%**.
- Implemented full-stack **observability** using **Grafana**, **Loki**, **Prometheus**, and **OpenTelemetry**, enabling proactive debugging, SLO tracking, and performance tuning across services.
- **Led SOC 2 Type 1 certification** as **SPOC**, completing the process in **1 month**. Trained our team on secure coding practices, and conducted code reviews focused on compliance and mentorship.
- **Tech Stack**: Node.js, Golang, FastAPI, NestJS, Next.js, Chakra UI, **Langchain**, **AI SDK**, **Mastra**, **OpenAI/Claude**, **Temporal.io**, Docker, KafkaJS, AWS, MongoDB, OpenTelemetry, PostgreSQL.

Software Engineer

Fountane Inc.

Minneapolis, Minnesota

Dec 2022 – May 2024

- Took full ownership of the **Parkpoolr** project, including **direct client engagement** and **leading an intern**, driving **full-stack development**, **infrastructure optimization**, and **bug resolution**. **Reduced GCP costs by 30%** and cut **bug fixing time** from **~30% of development effort** to a minimal fraction through **structured workflows** and **proactive diagnostics**.
- Contributed as a **Full Stack and Backend Engineer** on two additional projects, working with **Angular**, **React**, **Golang**, and **Node.js** to develop scalable services, build frontend features, and optimize **GraphQL/Hasura API** performance.
- Led migration from **VM-based deployments** to **Docker Swarm**, implementing **CI/CD automation** to improve **scalability** and **deployment efficiency**. Also integrated **SSO** and **role-based access control** with **Keycloak** to secure observability dashboards and internal monitoring systems.
- Developed a complete **observability stack** with **Grafana**, **Loki**, **Prometheus**, and **OpenTelemetry**, significantly improving **debugging**, **QA workflows**, and enabling **A/B testing analytics**. Also created an internal **AI Agent** (“**Mondai**”) for Slack, integrating with **Asana**, **Google Calendar**, **Keka HR**, and **Gmail** to unify **task tracking**, **scheduling**, and **HR insights** into a single **conversational interface** for internal team productivity.

Full Stack Developer (Intern to Engineer)

Pratilipi, 1DIGITALSTACK.AI, MADPAT GmbH

Bengaluru, South Delhi,

Germany

Jan 2021 – Nov 2022

- Developed and maintained **full-stack features** using **Vue.js**, **React.js**, **AngularJS**, **Node.js**, **Python**, **PostgreSQL**, **MongoDB**, and **Redis**, contributing to **production-scale applications** in publishing, analytics, and sales automation domains.
- Contributed to backend **infrastructure upgrades**, including migrating from **Node.js v6 to v14**, shifting from **MEAN.io to Express.js**, and building secure microservices using **RabbitMQ**, **OpenAPI**, and inter-process communication patterns.
- Helped build a secure **Credential Vault** using **double symmetric encryption** for storing **sensitive client data**; implemented **authentication** and **authorization** flows using **Keycloak**, **OAuth2.0**, **JWT**, and **Passport.js**.
- Designed data instrumentation and **analytics pipelines** with **Amplitude** for **A/B testing**, experiments, and **product metrics**, contributing to a **160% increase** in **web-to-app installs**.
- Automated development and onboarding environments using **Docker**, **.devcontainer**, and **Bash scripts** to speed up local setup and maintain **consistency across teams**.

Projects

Quantum Agent – Enterprise AI Automation

PersanaAI (2024)

- Production **AI Agent** built at **PersanaAI**, handling **20M+ requests** with external integrations, **NLP automation**, and fault-tolerant orchestration. (See “**Experience**” section for full details.)

C++ REPL – Interactive Language Runtime

c-repl/c-repl [🔗](#) [🔗](#)

- Developed a fully functional **C++ interpreter** with both **CLI** and **web REPL** interfaces, enabling interactive **C/C++ code execution** and **debugging**.
- Built **custom parsers** using **PEG.js**, interfacing with **GCC** and **GDB** for **real-time compilation** and **step-debugging**.
- Implemented **multi-process architecture** and **IPC**, along with a web-based frontend using **React.js**.
- **Tools Used:** Node.js, TypeScript, React.js, PEG.js, GCC, GDB

TCP Proxy Server

tcp-proxy [🔗](#) [🔗](#)

- Created a **transparent TCP Proxy Server** in **Node.js** enabling **client authentication** and **real-time connection logging**.
- Focused on **low-level socket handling** with extensibility for **debugging** and **network security** use cases.
- **Tools Used:** Node.js, Net module, TCP/IP stack

Technologies

Languages: TypeScript, Python, Golang, C++, SQL

Frontend: React.js, Angular, Next.js, Chakra UI, TailwindCSS

Backend: Node.js, FastAPI, NestJS, Express.js, GraphQL, Hasura, OAuth2.0, JWT, REST, RabbitMQ

AI/LLM Tooling: Langchain, AI SDK, OpenAI/Claude, Mastra

Infrastructure: Docker, GitLab CI/CD, Redis, KafkaJS, Temporal.io, AWS, GCP, OpenTelemetry (Grafana Stack, Jaeger)

Databases: PostgreSQL, MongoDB

Education

Dr. A. P. J. Abdul Kalam Technical University

Aug 2017 – May 2021

B.Tech in Computer Science and Engineering, Lucknow, Uttar Pradesh

- **Coursework:** DSA, OS, CN, DBMS, Software Engineering, AI, Web Technologies