

# SUYASH SHRESTHA

+977-9841309522 • Kathmandu, Nepal • [sthasuyash11@gmail.com](mailto:sthasuyash11@gmail.com)  
[suyashshrestha.com.np](http://suyashshrestha.com.np) • [github.com/sthsuyash](https://github.com/sthsuyash) • [linkedin.com/in/sthsuyash](https://linkedin.com/in/sthsuyash)

Software Engineer specializing in backend infrastructure, full-stack web applications and practical ML/AI integrations. Experienced in building reliable services, data pipelines, optimizing database schemas and developing RAG systems along with custom machine learning models to solve complex software challenges.

## TECHNICAL SKILLS

---

**Languages:** Python, JavaScript, Typescript, Golang

**Backend & Cloud:** Node.js, FastAPI, Docker, AWS

**AI/ML & Data:** RAG, Vector Databases, Elasticsearch, LLM agents, Pandas, NLP

**Frontend & Web:** React.js, Next.js, Playwright, Puppeteer

**Database & Tools:** PostgreSQL, MySQL, MongoDB, Redis, Firebase, Git, Linux

## EXPERIENCE

---

### Data Analyst & Backend Engineer

06/2025 - Present

*Open My Network* | New York, United States | (Part-Time)

- Migrated legacy reporting backend to Hexagonal Architecture, decoupling core business logic from database and external adapters for maintainability and clean testing.
- Built an automated PDF report generation pipeline using Puppeteer and AWS S3, converting dynamic HTML/CSS templates into high-fidelity reports at scale generating.
- Optimized legacy SQL database queries and data ingestion pipelines, significantly reducing query execution times and report generation bottlenecks.
- Containerized services with Docker and deployed to AWS, establishing production container hosting and automated end-to-end testing with Playwright.

### Software Engineer

09/2024 - Present

*Workharu Labs* | Kathmandu, Nepal

- Architected and launched a self-serve QR pay-to-print platform, building workflows for instant PDF conversion, mobile payments and CUPS-based IoT print server integration.
- Overhauled legacy RAG pipelines and vector search architecture, improving context retrieval quality and implementing automated evaluation frameworks to benchmark response accuracy.
- Orchestrated end-to-end AI workflows using local stacks, building resilient, stateful services for complex multi-step data processing and LLM evaluations.
- Built scalable web scraping and data ingestion pipelines, leveraging automated extraction and processing of external datasets to feed downstream AI models.
- Designed and deployed applications on AWS, managing cloud resources, database schemas and Playwright QA automation to ensure reliable production deployments.
- Engineered end-to-end blockchain systems, developing full-stack Web3 features and FastAPI backends to support dynamic token minting, buying and selling operations.

### ML/AI Engineer Intern

05/2024 - 08/2024

*Inspiring Lab* | Lalitpur, Nepal | Internship

- Designed custom document processing strategies for complex tabular data, solving text extraction and chunking issues in English and Devanagari /Nepali PDFs to significantly improve retrieval accuracy over standard parsers.
- Built a secure sandboxed execution environment for an autonomous data analysis agent, enabling it to safely run LLM-generated Python code for dynamic pandas analysis, plotting and report generation.

- Optimized vector search and embedding workflows using Elasticsearch, tuning hybrid retrieval and index settings to improve context relevance and retrieval speed.
- Developed an end-to-end RAG chatbot, integrating custom prompt engineering, context window management and reranking to cut hallucinations on domain-specific QA tasks.

### Full Stack Software Engineer

02/2023 - 03/2024

*Yakana IT Solutions* | Melbourne, Australia

- Built core features for multi-vendor web applications, including integrating third-party KYC APIs to handle user and merchant verification.
- Engineered backend services and interactive dashboards for real-time tracking systems, maintaining low-latency data sync across active sessions.
- Worked on full-stack web projects building dynamic UI components and handling real-time data flows.
- Refactored legacy backend query logic and frontend components under senior engineer guidance, optimizing database query performance and improving application response times to milliseconds.

### Software Engineer Intern

05/2022 - 11/2022

*Kiko Solutions* | Kathmandu, Nepal | Internship

- Developed internal and client-facing web application features and API endpoints in collaboration with cross-functional team members.
- Designed relational SQL database schemas and authored internal technical documentation to support backend maintainability.

## PERSONAL PROJECTS

---

- Dwello: *Full-stack property valuation platform with automated price estimation models*
- SajhaKhabar: *Automated news aggregation portal featuring NLP classification, sentiment analysis, summarization for Devanagari text.*
- DocuLume: *Multimodal RAG platform with custom PDF layout parsing and hybrid vector search.*
- Football Analysis System: *Computer vision pipeline tracking player movement and spatial match tactics.*
- AI/NLP TV Series Analysis System & Multimodal Sentiment Analysis System: *NLP platform featuring zero-shot classification, character networks, and persona chatbots.*

## EDUCATION

---

**Bachelor of Science in Computer Science and Information Technology** | Prime College

2021 - 2025

*(Affiliated to Tribhuvan University, Kathmandu, Nepal)*

- Member, Prime IT Club

## CERTIFICATIONS

---

- IBM Full Stack Software Developer Specialization | *Coursera*
- AWS Academy Graduate - AWS Academy Cloud Foundations | *AWS Academy*
- Machine Learning Specialization | *DeepLearning.AI / Stanford, Coursera*