Sabin Shrestha

sabin.shrestha.er@gmail.com +977 9848920281 Kathmandu, Nepal
linkedin.com/in/sabinshresthaa github.com/masabinhok sabinshrestha69.com.np

Education

IOE Pulchowk Campus, Tribhuvan University **Expected Graduation: 2027**
Bachelor of Engineering in Computer Engineering *Lalitpur, Nepal*

Certifications

AWS Certified Solutions Architect – Associate *Amazon Web Services*

Projects

Hamsa | *Source Code* **Python | FastAPI | Qdrant | SigLIP | Docker | Next.js**
Open-source fact-check triage for Nepali disaster misinformation

- Turned suspicious posts and video frames into **cited verdicts** via a **6-stage pipeline** (ingestion through deployment) on **5 open-weight models**, with **zero proprietary AI APIs**
- Flagged recycled media by pairing **perceptual hashing** with **SigLIP embeddings**, and enabled cross-lingual **Nepali/English RAG** retrieval over a **Qdrant** vector index
- Kept a human in the loop with a **4-outcome router** (merge duplicate, fast-track, human review, investigate), so no verdict publishes without a fact-checker's approval
- Made quality measurable with an **evaluation harness** over labelled debunked and genuine posts; shipped with **Docker Compose**, architecture docs and decision records across **133 commits**

Sarathi | *Source Code* **Python | LangGraph | Ollama | Chroma | FastAPI | Next.js**
Fully local RAG assistant for IOE entrance and admission queries (ANLP minor project)

- Answered admission, fee, cutoff and deadline queries with **no cloud inference** via a **LangGraph** flow (rewrite → retrieve → exact lookup → answer) over **Qwen2.5-7B** and **BGE-M3** on **Ollama**
- Built the ingestion layer: scrapers for **7 official notice boards**, PDF extraction with legacy **Preeti-font decoding**, and a **human-approval queue** before any text reaches the **Chroma** index
- Eliminated digit drift by routing form-number and merit-rank queries to an **exact lookup over a 7,179-row pass list**, not vector search; re-ranking plus a relevance floor makes the bot decline rather than guess
- Pre-resolved **Bikram Sambat** dates to AD before generation and served **SSE-streamed** chat, notices and token-guarded admin APIs through **FastAPI**; one-command **Docker Compose** setup

sms-api | *Source Code* **NestJS | Kafka | PostgreSQL | Prisma | Docker**
Event-driven backend for a school management system

- Decoupled **7 NestJS microservices** (gateway, auth, student, teacher, academics, activity, email) through event-driven messaging on a **3-broker Apache Kafka** cluster
- Enforced data ownership with a **database-per-service** design across **6 PostgreSQL databases**, managed through **Prisma** schemas, migrations and seed scripts
- Secured **4 user roles** (admin, teacher, student, guardian) with **JWT** access/refresh tokens and **RBAC**; documented every service with **Swagger**
- Reduced reviewer onboarding to a **single command** by containerizing the full stack with **Docker Compose**

Technical Skills

Languages: Python, SQL, TypeScript, JavaScript
Data & AI: Data pipelines, web scraping, PDF extraction, RAG, embeddings, LangGraph, Ollama, evals
Data Stores: PostgreSQL, Prisma, Qdrant, Chroma
Backend & Streaming: Apache Kafka, FastAPI, NestJS, REST, SSE, JWT/RBAC, Swagger
Cloud & Tools: AWS, Docker, Docker Compose, Git/GitHub, Next.js, uv