

AAGAMAN BHATTARAI

AI/ML Engineer

+977-9823763462 • aagamanbhattarai@gmail.com • <https://www.linkedin.com/in/aagaman-bhattarai-07068122a/>
• aagamanbhattarai.com.np • Lalitpur, Nepal

Summary

AI/ML Engineer with hands-on experience building production-oriented machine learning, Generative AI, RAG, text-to-SQL, and analytics automation systems across structured and unstructured data. Experienced in Python, FastAPI, SQL, data pipelines, retrieval systems, LLM evaluation, prompt engineering, agentic workflows, and model validation. Strong background in building AI systems that support forecasting, reporting, decision-making, and business intelligence. Experienced with LLM benchmarking, chunking strategy evaluation, multilingual English/Nepali RAG experiments, and local sub-10B model evaluation.

Experience

Wiseyak Inc Kathmandu, Nepal
AI/ML Engineer 11/2025 - Present

- Built an LLM-powered conversational analytics system enabling users to query structured and unstructured business datasets through natural language and voice in English, Nepali, Maithili.
- Developed a two-way AI voice call system where users can ask questions over a call and receive human-like spoken responses generated from retrieved business context.
- Designed RAG, data ingestion, and retrieval pipelines to ground AI responses in internal documents, database schemas, and business data.
- Implemented validation, testing, and backend workflows using Python, FastAPI, SQL-based systems, and production-oriented AI engineering practices.

Ujyalo Ventures Kathmandu, Nepal
Associate AI Engineer 02/2025 - 10/2025

- Built and deployed a production text-to-SQL system for querying internal databases using natural language.
- Reduced manual reporting workload by 40%, enabling faster operational decision-making across teams.
- Developed SQL-based data pipelines to support business intelligence dashboards, analytics workflows, and automated reporting.
- Performed data profiling, transformation, validation, and anomaly checks to improve the reliability of downstream analytics.
- Implemented prompt engineering, SQL generation logic, schema grounding, and fallback mechanisms to reduce incorrect query outputs.
- Worked with structured business datasets and translated business requirements into analytical workflows and AI-powered solutions.

Skills

Generative AI & LLM Systems: LLMs, RAG, AI Safety, text-to-SQL, agentic workflows, LLM benchmarking, LangChain, LanGraph

Machine Learning / AI: PyTorch, Scikit-learn, Pandas, ML pipelines, feature engineering, clustering, forecasting models, Etc.

AI & Tools: FastAPI, Docker, Git, Azure, REST APIs, MCP, Agentic Coding (Codex, ClaudeCode)

Projects

AI for Business Insights – Text-to-SQL Analytics System 04/2025

LlamaIndex, Agent Driven Architecture, FastAPI, SQLAlchemy, Fuzzy Matching, Sandboxing, Azure OpenAI)

- Built a production-ready text-to-SQL analytics system that converts natural-language questions into validated SQL queries and business insights.
- Designed schema retrieval, fuzzy matching, SQL validation, fallback handling, and report-generation workflows to improve answer reliability.

LLM & RAG Benchmarking – English and Nepali 01/2026

Chunking strategies comparison, Benchmarking on Medical & Financial Documents, LLM Guardrails and Jailbreaking

- Benchmarked multiple local sub-10B LLMs across English and Nepali tasks for answer quality, latency, reliability, and deployment feasibility.
- Evaluated chunking strategies including recursive, section-aware, parent-child, paragraph-based, and table-aware chunking for RAG pipelines.
- Evaluated using hit rate@k, evidence recall@k, faithfulness, answer relevance, latency, and duplicate chunk rate.

Education

Kathmandu University Dhulikhel, Nepal
Bachelor of Technology, Artificial Intelligence | GPA: 3.79 / 4.0 2021 - 2025