Utsav Manandhar

+977-9823287359| utsavmanandharz@gmail.com | linkedin.com/in/utsav-manandhar |
github.com/Utsav-Manandhar

Software Engineer specializing in AI/ML and Backend systems. Experience with Go, Python, and RAG pipelines. Hands-on experience in building scalable APIs and fine-tuning LLMs for production environments

Technical Skills

Languages: Python, C/C++, Go, SQL, JavaScript, Bash/Shell Scripting
AI/ML & LLMs: PyTorch, LangChain, Transformers (Hugging Face), RAG (Retrieval-Augmented Generation), LLM Orchestration, Scikit-learn
Data Engineering & Databases: Pandas, NumPy, Snowflake, Neo4j (Graph), Qdrant (Vector), PostgreSQL, Redis
Cloud & Devops: AWS, Google Cloud Platform (GCP), Docker, NGINX, Prometheus, Grafana, Git/GitHub, Linux
Web & Backend: FastAPI, Django, React, Next.js, Tailwind CSS

Experience

L1 Software Engineer (AI/ML)

Vivasoft Nepal

Lazimpat, Kathmandu

Dec 2025 - Present

• Maintain and extend a large-scale backend codebase in Go and Python, working across services, data layers, and internal APIs to ensure performance, reliability, and clean service boundaries.
• Worked with cloud deployment and DevOps workflows (CI/CD, containers, and service configuration) to deploy and operate backend services reliably in production.

Associate Software Engineer (AI/ML)

Vivasoft Nepal

Lazimpat, Kathmandu

Jul 2025 - Dec 2025

• Fine-tuned speech-to-text (STT) models for low-resource languages, handling data preprocessing, tokenization, and training workflows to improve transcription accuracy.
• Worked on in-house AI projects focused on Retrieval-Augmented Generation (RAG), integrating vector search and retrieval pipelines to ground model responses in domain-specific data.
• Worked on audio-focused AI systems, including speech preprocessing, feature extraction, and model experimentation, supporting multiple audio understanding and generation use cases.

Backend/AI Developer Intern

Uniseek Education Services

Anamnagar, Kathmandu

Jun 2025 - Jul 2025

• Developed AI agent–based learning tools by integrating agent logic, prompt workflows, and backend services in Python.

• Analyzed and optimized backend delivery architecture using Cloudflare (caching rules, CDN), drastically reducing page load latency and improving overall application performance.
• Designed, maintained, and extended RESTful APIs for the backend codebase using Python and Django, ensuring reliability and scalability.

Training and Certifications

AI Fellowship - Fusemachines   April 2025

• Built and fine-tuned machine learning models using Python libraries such as scikit-learn, TensorFlow, and Pandas.
• Completed assignments and gained hands-on experience with concepts like supervised learning, neural networks, and model interpretability.

Data Warehousing and Business Intelligence — LIS Nepal Pvt. Ltd.   Feb 2024

• Developed a strong understanding of data warehousing concepts, including ETL processes, dimensional modeling, and star/snowflake schemas using **Snowflake**.
• Gained proficiency in both basic and advanced SQL concepts, with strong expertise in writing complex queries.
• Worked with data warehousing and data visualization tools such as **Snowflake** and **Power BI.**

AWS Academy Cloud Foundations — AWS Academy   Aug 2024

• Designed and configured secure cloud infrastructures using AWS IAM, Virtual Private Cloud (VPC), and EC2, ensuring robust access control and network isolation.
• Deployed and managed scalable web applications using AWS Elastic Beanstalk, gaining expertise in cloud-based application lifecycle management.
• Implemented serverless solutions with AWS Lambda, optimizing compute resources and reducing operational costs through event-driven architecture

Projects

VivaDai — Memory Management Chat System & Voice Agent

• Designed and developed an organizational memory platform featuring real-time Speech-to-Speech (STS) pipelines, enabling fluid voice interactions using FastRTC.
• Architected a RAG (Retrieval-Augmented Generation) system by combining Neo4j (knowledge graph) and Qdrant (vector search) to structure and retrieve institutional knowledge with high precision.
• Built high-performance backend services with FastAPI to orchestrate the data flow between audio processing, vector retrieval, and Gemini LLM responses.
• Implemented adaptive memory mechanisms that capture and contextualize team interactions, allowing the agent to provide intelligent, context-aware answers.

Kothon Analytics Dashboard — End-to-End AI Voice Agent & Analytics Platform
• Re-architected and maintained a scalable microservices backend using Go and Python, overseeing the database schema changes and API design lifecycle.
• Managed end-to-end deployment and operations on Google Cloud Platform (GCP), ensuring production stability for the cloud-hosted voice agent system.
• Implemented complex business logic and analytical features to track telecom metrics, such as churn risk and customer engagement, derived from AI-driven voice interactions.

Aakashwani & Dogmatix — AI Voice Agent Incoming/Outgoing Platform
• Engineered a full-stack AI telephony platform, building both the FastAPI backend services and a Jinja2-based administrative dashboard for campaign management.
• Bridged traditional telephony with modern AI by connecting PJSIP and Asterisk to the Gemini LLM via WebSockets, enabling low-latency, real-time voice conversations.
• Orchestrated the complete call lifecycle, handling inbound/outbound routing, AI processing, and system administration in a unified solution.

AI-based Vocal Isolation and Separation Using CNN model

• AI- based Vocal isolation from songs using Convolutional Neural Networks
• Prepared and handled dataset of audio files using Python and Librosa.
• Designed and fine tuned an AI model in Tensorflow to isolate vocals from musical tracks.

Bayesian Network for Weather Prediction

• Bayesian Network based weather predictor
• Prepared and handled dataset of weather reports using Python and Kaggle
• Designed an AI model to make predictions based on that data.

Education

Pulchowk Campus , Tribhuvan University   Pulchowk, Lalitpur

Bachelor of Engineering in Computer Engineering | **First Division**   May 2021 - April 2025

Nepal Mega College   Babar Mahal, Kathmandu

+2/SLC in Science (Physics) | GPA **3.71**   June 2018 - October 2020