AAYUSH SHAH KANU

Software Engineer
aayushshahkanu@gmail.com | +9779817326587 | Hattiban, LA road
Portfolio | **GitHub** | **Linkedin**

EDUCATION

Pulchowk Engineering Campus   Lalitpur, Nepal
Computer Engineering Bachelors

EXPERIENCE

Docsumo | Machine Learning Engineer   Singapore | Dec 2022 - Present
- Optimized legacy code to increase extraction Straight-Through Processing (STP) rates from 75% to over 90%, with peak performance reaching 95% for select projects
- Co-led a time-critical project integrating Natural Language Processing (NLP) and Computer Vision (CV), achieving 99% accuracy
- Developed and implemented a Large Language Model (LLM) API, enabling faster document extraction demos for sales and customer support teams by skipping data collection time
- Managed end-to-end projects from requirements gathering to client handover, ensuring stakeholder satisfaction and on-time delivery
- Conducted experiments on Machine Learning (ML) models, testing extraction hypothesis, improving accuracy and resolving issues.
- Engineered a synthetic data generation pipeline, increasing data creation efficiency for fixed/structured forms
- Refactored an Insurance Form Extraction API processing 1.5M+ monthly requests, reducing per-document extraction time by 20% and improving accuracy by 8%.
- Oversaw APIs for top-tier clients, collaborating with Customer Success and Product teams to resolve issues and implement new features
- Built data pipelines and designed a comprehensive monitoring and reporting system for user data analysis, supporting Product, CS, ML, and Marketing teams to meet specific use cases.
- Collaborated cross-functionally to align data solutions with business goals.
- Streamlined PQL interactions, improving user engagement and conversion efficiency.

Stealth Startup | Founding Engineer/Freelancer   Remote | Aug 2024 - November 2024
- Developed full-stack carbon accounting platform, including a customer-facing landing site, responsive web application, and RESTful backend API
- Engineered automated CI/CD pipelines for backend infrastructure using Google Cloud Platform (GCP)
- Optimized frontend deployment workflows via Vercel-integrated CI/CD processes
- Drove product strategy by conducting competitive market analysis to define MVP scope, prioritizing high-impact features aligned with sustainability compliance standards.

Bajra Technologies LLC | Software Engineer Trainee   New York, US | Sep 2022 - Dec 2022
- Worked with Python, JavaScript, XML, PostgreSQL, and Odoo to create ERP solutions.
- Developed health and productivity projects, including Eye Health Promotion App and Task Management App.
- Worked closely with a team to learn best practices in software development and deployment.
- Created and delivered presentation on different technologies and tech stack

Fusemachines.AI | AI Fellow   New York, US | Mar 2022 - Jan 2022
- Completed Microdegree on Machine Learning and Deep Learning.
- Learned to leverage different techniques and models to understand data.
- Developed an App to showcase development of end-to-end Machine Learning and software pipeline.

esatya(Rumsan Group) | Blockchain Fellow   Kathmandu | Apr 2022 - Jul 2022
- Completed the fellowship on Blockchain in ethereum ecosystem
- Learned about concept of blockchain, Ethereum ecosystem, dapp and its development and deployment
- Developed a Voting App to showcase development and deployment of Ethereum based DApp

SKILLS

Programming Languages:	Python, Javascript
Libraries/Frameworks:	Flask, FastAPI, Pytorch, Pandas, scikit-learn, numpy, transformers, React Native, mlflow, openai, React
Tools / Platforms:	Docker, Postman, GCP, Git
Databases:	PostgreSQL, MongoDB

PROJECTS / OPEN-SOURCE

Docbite: Information Extraction from Structured Documents | Link
Python,Huggingface,React,Mongodb
- Utilized ML algorithms for accurate data extraction from structured documents such as Tax Forms, Insurance Forms.
- Developed annotation tool for efficient data preparation.
- Created training module for continuous improvement of data extraction accuracy through annotated data.
- Designed flexible platform for API integration, streamlining document processing workflows.

Agent Engine | Link   *Python*
- Built a Python framework for dynamic, AI-driven workflows with multi-LLM integration.
- Designed customizable agents, task routing, and stateful memory for adaptive systems.
- Enabled runtime task generation and visual workflow design via YAML/JSON or Python API.
- Developed modular agent architectures and pre-built templates for streamlined workflows.

Chatterbox | Link   *python, streamlit*
- Developed Chatterbox, a RAG chatbot for document interaction, enhancing retrieval efficiency.
- Enabled session storage to retain context over long periods, allowing users to continue previous conversations.

News LLM | Link   *Python, Openai, OpenProvider*
- Developed a tool to scrape and summarize news articles using LLMs.
- Designed an extendable architecture for easy addition of new scrapers
- Integrated real-time news delivery to Slack and other communication channels.
- Designed a modular system for easy extension with custom scrapers, summarizers and communication channels.

CERTIFICATIONS

- Deep Learning Specialization - **Coursera**
- DeepLearning.AI TensorFlow Developer Specialization - **Coursera**
- AI for Medicine Specialization - **Coursera**
- Build Basic Generative Adversarial Networks (GANs) - **Coursera**
- Natural Language Processing with Classification and Vector Spaces - **Coursera**
- AWS Academy Graduate - AWS Academy Machine Learning Foundations - **AWS**

HONORS & AWARDS

- Ncell Scholarship & Excellence Award