Sadiksha Chitrakar

Miami, FL | (203) 927-6288 | sadiksha.chitrakar101@gmail.com | linkedin.com/in/sadikshachitrakar | sadikshachitrakar.com.np

Summary

AI Engineer with an M.S. in Data Science and 2+ years of software engineering experience building end-to-end LLM and ML systems, from RAG and agentic workflows to LLM evaluation, guardrails, APIs, and cloud deployment.

Education

Master of Science in Data Science — August 2024 - May 2026 | GPA: 3.97
University of New Haven — West Haven, Connecticut, United States

Bachelor of Science (with Honors) in Computer Science — October 2019 - November 2022 | GPA: 3.94
Coventry University — Coventry, England, United Kingdom

Experience

AI Engineer — August 2026 - Present
AI Trusted Advisors — Las Vegas, Nevada, United States (Remote)

- Architected and deployed an explainable AI career recommendation platform (Python, FastAPI, React, PostgreSQL, AWS) pairing ML ranking with LLM reasoning to deliver transparent recommendations and automated skill-gap analysis.
- Built an autonomous content agent (Groq, GPT-OSS-120B) publishing 4 articles daily across 199 topics in 17 categories, with weighted rotation and balance tracking to prevent duplicates; grew indexed content 30% and organic traffic 40%.
- Developed an automated LLM evaluation layer (Pydantic schema validation, section and keyword-relevance checks, brand-leakage guardrails) with regenerate-on-failure, backoff retries, and atomic writes that protect publishing state.

Graduate Research Assistant — August 2025 - December 2025
University of New Haven — West Haven, Connecticut, United States

- Built and evaluated end-to-end ML pipelines for biological age prediction on 75,000+ NHANES records, achieving a 3.5-year MAE ($R^2 = 0.95$) under 5-fold cross-validation, with XGBoost reducing MAE by 18% versus Random Forest.
- Engineered reproducible data pipelines across multiple NHANES survey cycles using MICE imputation, data standardization, feature engineering, cohort construction, and feature-importance analysis to identify predictive biomarkers.

Software Engineer — January 2024 - January 2025
Xuno — Ellensburg, Washington, United States

- Architected and shipped the Xuno cross-border payments app on iOS and Android, structuring it into modular money-transfer, KYC, and wallet flows; the app reached a 4.7-star rating and 500+ installs.
- Reduced post-deployment bugs by 30% with unit, widget, and integration tests prioritized from production logs and metrics.

Flutter Developer — November 2022 - January 2024
Code Himalaya — Lalitpur District, Nepal

- Developed and maintained cross-platform mobile apps integrating REST APIs, and cut release-cycle time by 70% by automating build and deployment with Fastlane CI/CD, publishing 30+ apps to the App Store and Google Play.

Projects

CampusGuide: Agentic RAG University Assistant — GitHub

- Built a RAG system over university documents (PDF, DOCX, HTML) using LangChain and ChromaDB, and exposed campus tools (document search, office/program/deadline lookup, deadline calculator) via an MCP server and FastAPI.
- Benchmarked Llama-3-8B, Mistral-7B, and Phi-3-Mini with an automated LLM-as-a-Judge harness (PatronusAI Lynx-8B); Llama-3-8B led with 53.2% factual accuracy and 0.54 grounding, versus 42.7% and 43.6%.
- Reduced hallucinations on out-of-scope queries by engineering an agentic LangGraph workflow (query routing, tool calling, context grading, rewrite-and-retry, verification, escalation) that enforced context-grounded responses and correct abstention where weaker models fabricated content.

AWS Serverless Data Pipeline for Demand Forecasting — GitHub

- Built a serverless AWS ETL pipeline that automatically cleans and loads 100,000 multi-store retail records from S3 into PostgreSQL using Lambda, with Glue and Athena enabling SQL analytics.
- Trained Random Forest demand-forecasting models per product category on SageMaker and exposed predictions through a REST API (API Gateway, Lambda) powering Power BI dashboards for non-technical stakeholders.

Keypoint Detection for Building Components — GitHub

- Fine-tuned a Keypoint R-CNN (ResNet-50/101-FPN) in PyTorch, comparing both backbones, to detect building components and regress semantic keypoints on a custom COCO-format dataset (200 images augmented to 1,000 with Albumentations).
- Achieved 0.89 mAP and 0.82 keypoint mAP (OKS), cutting validation loss from 10.05 to 3.73.

Skills

Languages & Frameworks: Python, SQL, JavaScript, TypeScript, React, Flutter, Dart, R
LLMs & GenAI: RAG, GraphRAG, Agentic AI, LangChain, LangGraph, MCP, LLM Evaluation, Hugging Face, ChromaDB
Machine Learning: PyTorch, TensorFlow, scikit-learn, XGBoost, NLP, Computer Vision, Feature Engineering
Cloud & MLOps: AWS (S3, Lambda, Glue, Athena, SageMaker, RDS, API Gateway), Docker, FastAPI, CI/CD, Git, pytest
Data: Pandas, NumPy, Pydantic, ETL, Data Modeling, Apache Spark, PostgreSQL, MySQL, MongoDB, Firebase

Certifications

Power BI (Data Modeling & Visualization) | Apache Spark Programming with Databricks | MLOps Essentials: Model Development & Integration