Shiven Paudyal

562-348-5148 | shiven19p@gmail.com | LinkedIn | GitHub | Portfolio | Long Beach, CA

Technical Skills

ML & Deep Learning: PyTorch, TensorFlow, scikit-learn, XGBoost, model training, fine-tuning, hyperparameter optimization, feature engineering, A/B testing
Generative AI & LLMs: LangChain (AgentExecutor, ReAct), RAG pipelines, FAISS, OpenAI API, Ollama, Llama 3.3 70B, prompt engineering, LLM-as-a-judge, tool orchestration
MLOps & Data: MLflow, EvidentlyAI, model versioning, data drift detection, CI/CD for ML, Python (Pandas, NumPy), SQL, PostgreSQL, ETL pipelines, REST APIs
Production & Infrastructure: AWS (SageMaker, Lambda, S3, CloudWatch), Docker, GitHub Actions, FastAPI, Cloudflare Workers, Git, GitHub Copilot

Experience

Zentrais LLC — Los Angeles, CA
Machine Learning Engineer Intern — Feb 2026 – May 2026

- Improved **model accuracy by 20%** by engineering end-to-end **model training pipelines** covering **data preprocessing**, **feature engineering**, and **hyperparameter tuning** validated through structured **A/B testing**.
- Reduced **model degradation incidents by 30%** by architecting **production monitoring pipelines** in **EvidentlyAI** and **MLflow** to continuously track **data drift**, accuracy, and quality across live inference environments.
- Accelerated **model release cycles by 25%** by containerizing **inference services** with **Docker** and automating **CI/CD deployments** on **AWS SageMaker** via **GitHub Actions**, maintaining production-grade **Python** codebases.

UiPath — Chennai, India
Robotic Process Automation Developer Intern — Jan 2023 – Jun 2023

- Cut **data quality issues by 35%** across 500K+ weekly records by engineering **data preprocessing** and **feature extraction** pipelines in **Python** and **SQL** to surface and remediate recurring failure patterns.
- Reduced **automation failure risk by 20%** by training and evaluating **predictive models**, translating business requirements into **model specifications** with cross-functional engineering and product teams.
- Slashed **manual processing time by 40%** by implementing **ETL pipelines** for automated data ingestion and transformation, improving **data pipeline** throughput for downstream **model training** workflows.
- Communicated **model performance findings** across **10+ workflows** via structured reports and dashboards, translating technical outputs into actionable insights for non-technical stakeholders.

Projects

Fraud Detection MLOps Pipeline | *Python, XGBoost, MLflow, FastAPI, Docker, AWS SageMaker*

- Achieved **92% model accuracy** in production by engineering a complete **MLOps lifecycle** on **AWS SageMaker**, managing 40+ experiment runs in **MLflow** with systematic **hyperparameter tuning** and **model versioning**.
- Prevented **accuracy degradation** below 5% by integrating **EvidentlyAI** for automated **data drift** detection, triggering **model retraining** and serving updated inference via a containerized **REST API**.

AI Career Coach Agent | *TypeScript, Cloudflare Workers AI, Durable Objects, Llama 3.3 70B*

- Deployed a production **LLM inference** system on **Cloudflare Workers** running **Llama 3.3 70B**, serving real-time **WebSocket** chat with **PDF resume parsing** and structured ATS feedback generation at edge latency.
- Engineered **persistent per-user memory** using **Durable Objects** with SQLite, maintaining full conversation context across sessions in a stateless **serverless** architecture – live at `cf-ai-career-coach.shiven19p.workers.dev`.

Job Tracker | *Python, FastAPI, Groq LLM, Microsoft Graph API, PostgreSQL, Next.js*

- Achieved **95% extraction accuracy** by building an **LLM inference pipeline** with **Groq** to parse structured fields from unstructured Outlook emails via **Microsoft Graph API** OAuth2 – deployed on Render with auto-scaling.
- Shipped a production **REST API** in **FastAPI** with real-time sync, manual review queue, and **PostgreSQL** persistence – fully deployed on Vercel and Render with **GitHub Actions CI/CD**.

Customer Churn Prediction | *Python, XGBoost, SHAP, MLflow, Scikit-learn, Pandas*

- Attained **94% AUC** by training an **XGBoost** classifier with **hyperparameter optimization** and **feature engineering**, tracking all runs in **MLflow** for reproducible **model versioning** and experiment comparison.
- Validated **feature importance** alignment with domain knowledge using **SHAP** explainability – surfacing contract type, tenure, and monthly charges as top drivers before production sign-off.

Education

California State University, Long Beach — Long Beach, CA
Master of Science in Computer Science, GPA: 3.77 — Aug 2024 – May 2026

SRM University — Chennai, India
Bachelor of Technology in Computer Science — Aug 2019 – May 2023

Certifications and Publications

AWS Certified AI Practitioner (2025) | **AWS Certified Cloud Practitioner** (2025) | **Machine Learning Specialization** – Stanford/Coursera
Sign Language Translation Using Deep Learning – Published, IRCICD'23