

Khushal Hemant Sharma

Pleasanton, CA | 623 275 8176 | khushalrs@gmail.com | [LinkedIn](#) | [GitHub](#) | [Website](#)

Summary

Machine Learning Engineer with hands-on experience building and optimizing ML models and GenAI applications. Skilled in Python, PyTorch, TensorFlow, scikit-learn and frameworks like LangChain for RAG and embeddings. Developed and deployed production-grade AI services and chatbots using Flask/FastAPI, Docker, CI/CD pipelines and Azure/AWS cloud infrastructure. Proven ability to preprocess complex datasets and collaborate with cross-functional teams.

Education

Arizona State University, Tempe

M.S., Robotics and Autonomous Systems (AI concentration)

Aug 2024 - May 2026

Tempe, AZ

SVKM Narsee Monjee Institute of Management and Studies (NMIMS)

B.Tech., Artificial Intelligence

Jun 2020 - May 2024

Maharashtra, India

Experience

SiClarity Software Solutions

Software Development Intern

Apr 2024 - Jun 2024

Maharashtra, India

- Implemented and integrated predictive ML models into three Flask-based backend services to enhance internal workflows.
- Engineered REST endpoints for a knowledge-retrieval chatbot, preprocessing log and document data to deliver median 200–300 ms response times.
- Streamlined ML service deployment by migrating CI/CD pipelines from GitLab to GitHub Actions with automated testing and containerized builds, boosting reliability and consistency.

Celebal Technologies

Data Science Intern

Jan 2024 - Apr 2024

Maharashtra, India

- Architected a RAG-based document Q&A chatbot using OpenAI APIs, LangChain, FAISS, and RedisSearch for semantic enterprise search.
- Automated ingestion and preprocessing of unstructured documents with Azure Document Intelligence, CosmosDB, and Blob Storage to support scalable AI search pipelines.
- Assessed and benchmarked FAISS versus RedisSearch for semantic retrieval, analyzing quality, latency, and scalability trade-offs.

Projects

SimuTrader - Portfolio Strategy Simulator

- Built Python/FastAPI APIs for strategy creation, asynchronous backtest execution, equity curves, trade logs, tax views, and portfolio comparison across multiple simulation runs.
- Designed reproducible run-tracking workflows with persisted metadata, configurable transaction costs, long/short strategy support, and US/India tax logic for realistic portfolio evaluation.
- Integrated Redis/Celery-based asynchronous task execution to handle heavier backtest workloads without blocking API responses, improving backend responsiveness and scalability.

Serverless Grocery Price Monitoring

- Built a serverless data ingestion pipeline tracking 800 products across 10 zip codes, storing time-series price history in DynamoDB to support alerts, trend analysis, and product watchlists.
- Developed AWS Lambda microservices for ingestion, product search, price history retrieval, alert evaluation, and API responses, enabling a scalable event-driven backend architecture.

Technical Skills

- **Languages:** Python, Java, SQL, TypeScript
- **AI / ML:** PyTorch, TensorFlow, scikit-learn, NLP, Deep Learning, Computer Vision, Model Evaluation, Machine Learning, Statistical Modeling, Logistic Regression, Time Series Analysis, Random Forests, Support Vector Machines, XGBoost, Recurrent Neural Networks, Credit Scoring, Fraud Detection, Financial Forecasting
- **GenAI / LLMs:** LangChain, OpenAI APIs, RAG, Embeddings, Prompt Engineering, FAISS, RedisSearch, Vector Databases
- **Backend / MLOps:** Flask, FastAPI, REST APIs, Microservices, Docker, GitHub Actions, CI/CD
- **Cloud/Data:** Azure Document Intelligence, CosmosDB, Blob Storage, AWS Lambda, API Gateway, DynamoDB, S3, CDK, PostgreSQL, DuckDB, Redis, Celery
- **Core CS:** Data Structures & Algorithms, OOP, SDLC, Design Patterns

Research & Publication

- Kulkarni, V., Sharma, K., Shah, M., & Kulkarni, A..LIDNeRF.*International Journal of Intelligent Information Technologies*, 21(1), 1–18. 2025. doi:10.4018/ijit.369336