

Vinith Kumar Gudibanda Prasannakumar

AI / ML Engineer | LLMs, RAG & Agentic AI | Machine Learning & NLP | MLOps | Python & AWS
089 244 1008 | vinithvkk050@gmail.com | github.com/Vinithvk98 | [Portfolio](#) | [Blog](#) | Dublin, Ireland

SUMMARY

AI/ML engineer with 5+ years designing, developing, and deploying production AI and machine learning systems in Python. Builds LLM applications, agentic workflows, and RAG solutions with vector search and embeddings, alongside classical ML and deep learning for prediction, classification, fraud and anomaly detection, and forecasting. Takes models from prototype to production with MLOps and LLMops, CI/CD, containerisation, monitoring, and evaluation on AWS, and applies security, governance, and responsible AI. MSc in Data Analytics (Machine Learning and AI), National College of Ireland, Dublin.

AREAS OF EXPERTISE

- **AI & LLMs:** LLM Applications | Agentic AI & Workflows | Retrieval-Augmented Generation (RAG) | Prompt & Context Optimisation | LangChain | LangGraph | MCP | OpenAI & Anthropic APIs
- **Machine Learning & NLP:** Supervised & Unsupervised ML | Deep Learning (PyTorch, TensorFlow) | NLP & Transformers | Fraud & Anomaly Detection | Forecasting | scikit-learn | XGBoost
- **MLOps & Cloud:** MLOps / LLMops | CI/CD | Docker | AWS (SageMaker, Lambda, S3, ECS) | Model Deployment | Monitoring, Evaluation & Observability
- **Engineering & Data:** Python (Advanced) | SQL | REST APIs | Data Pipelines | Vector Databases & Embeddings | AI Security & Responsible AI | Stakeholder Collaboration

EDUCATION

MSc in Data Analytics, National College of Ireland, Dublin

2023 - 2024

- Machine Learning and AI | Applied Predictive Analytics | Cloud Computing | Database Design | Data Mining and Statistical Analysis

WORK EXPERIENCE

Data and Insights Analyst, Nando's Ireland, Dublin

May 2024 - Present

- Monitor and analyse 5 operational dashboards daily covering sales, labour, cost, and customer metrics, tracking 10+ KPIs to surface performance trends and flag anomalies to management before they escalate.
- Identify 3 to 4 recurring performance gaps per quarter through structured trend and variance analysis, delivering targeted recommendations that contribute to measurable improvements in operational efficiency.
- Conduct weekly sales forecasting and actuals-vs-targets analysis across multiple revenue streams, using AI tools to accelerate analytical workflows and enable more accurate staffing and cost-control planning.
- Translate complex data patterns into concise, stakeholder-ready insights during weekly stand-ups, directly influencing operational and resourcing decisions with management and operations leads.

AI Engineer, Infidata Technologies, India

Jan 2021 - Dec 2022

- Built and deployed AI, machine learning, and NLP solutions on a platform processing 10M+ daily transactions, including fraud and anomaly detection, reducing false positives by 25%.
- Developed LLM and NLP workflows with embeddings, retrieval, and prompt optimisation, integrating AI via APIs and custom orchestration, and cutting operational overhead by 50%.
- Deployed production-ready AI and ML services with monitoring, testing, and evaluation frameworks measuring quality, accuracy, latency, and cost.
- Applied MLOps and LLMops with version control, CI/CD, and Docker, embedding security, governance, and responsible AI, and cutting release cycle time by 40%.
- Collaborated with software engineers, data engineers, and business stakeholders, and maintained architecture, integration, and deployment documentation.

AI / Python Engineer, Infidata Technologies, India

Jul 2019 - Dec 2020

- Built Python AI and ML services, REST APIs, and data pipelines on AWS within Agile squads, integrating AI capabilities into enterprise applications.
- Reduced critical query execution time by 35% by optimising SQL and indexing on high-volume databases supporting AI and ML workloads.
- Supported deployment, monitoring, and evaluation of production AI and ML services, building MLOps and observability fundamentals.
- Built embeddings and retrieval prototypes and API integrations connecting AI to enterprise applications and third-party services.

PROJECTS

AutoDS-Copilot (Multi-Agent AI Data Science System) | [GitHub](#)

Personal Project

Python | LangGraph Multi-Agent | MCP Server | RAG & Vector Store | Human-in-the-Loop | scikit-learn

- Built a no-code, agentic AI system running the full ML workflow end to end, with a LangGraph multi-agent graph, RAG-grounded recommendations with citations, human-in-the-loop approvals, and an MCP server exposing 9 tools, all leakage-safe and pytest-covered.

SKU Sentinel (ML Evaluation & Drift Monitoring) | [GitHub](#)

Personal Project

Python | PyTorch LSTM | 8-Model Benchmark | Drift & Anomaly Detection | Monitoring & Alerting

- Built an open-source ML evaluation and monitoring system benchmarking 8 models and detecting drift and anomalies at 86% recall on real data, with a calibration fix that cut false alarms from 170 to 5.

NextMove (Prescriptive Decisioning) | [GitHub](#)

Personal Project

Python | Forecasting | Newsvendor Optimisation | Uncertainty Modelling | Backtesting

- Built a prescriptive system turning forecasts and uncertainty into optimised ordering decisions using newsvendor economics, adding \$100k over naive policies in backtesting.