Deepanshu Dariya

Guildford, Surrey, UK | +44 7721 557314 | deepanshudariya128@gmail.com
linkedin.com/in/deepanshu-dariya-424a64227 | github.com/dariyaDeepanshu | deepanshudariya.com | Eligible to work in the UK

PROFESSIONAL SUMMARY

MSc Artificial Intelligence candidate at the University of Surrey with a B.Tech in Computer Science (AI & Machine Learning) and hands-on experience building, evaluating, and deploying machine learning, deep learning, and large language model (LLM) systems. Proficient in Python, PyTorch, TensorFlow, and scikit-learn, with delivered projects across computer vision (Vision Transformers, CNNs), NLP (BERT/RoBERTa), retrieval-augmented generation (RAG), and agentic AI. Strong in model evaluation and MLOps practice (experiment tracking, drift monitoring, CI/CD), and in turning data into measurable outcomes. Seeking Machine Learning Engineer, AI Engineer, or Data Scientist roles.

EDUCATION

MSc, Artificial Intelligence | University of Surrey, Guildford, UK | Feb 2026 - Feb 2028 (expected)

- Relevant coursework: Advanced Machine Learning, Deep Learning, Natural Language Processing, Computer Vision, AI Planning, Data Science.
- Dissertation: Agentic AI system for detecting and explaining model drift in medical imaging.

B.Tech, Computer Science & Engineering (specialisation in AI & Machine Learning) | Vellore Institute of Technology (VIT-AP), Andhra Pradesh, India | Sep 2021 - Sep 2025

- CGPA: 7.68 / 10. Coursework in Machine Learning, Deep Learning, Data Structures & Algorithms, Databases, Statistics, and Computer Vision.

TECHNICAL SKILLS

Programming & Databases: Python, SQL, Java, R, Bash, PostgreSQL, SQLite, DuckDB, Neo4j

Machine Learning & Deep Learning: PyTorch, TensorFlow, Keras, scikit-learn, Hugging Face Transformers, XGBoost, LightGBM, CatBoost, supervised & unsupervised learning, model evaluation, hyperparameter tuning

NLP & Large Language Models: Transformers (BERT, RoBERTa), LLMs, Retrieval-Augmented Generation (RAG), prompt engineering, LangChain, sentence-transformers, text-to-SQL, agentic / multi-agent systems, Model Context Protocol (MCP)

Computer Vision: Convolutional Neural Networks (CNN), Vision Transformers (ViT), VideoMAE, TimeSformer, ResNet, VGG, image & video classification, action recognition

Data Science & Analytics: Pandas, NumPy, Matplotlib, exploratory data analysis (EDA), feature engineering, statistical analysis, data visualisation, experiment design

MLOps & Deployment: MLflow, Docker, Git, GitHub Actions (CI/CD), FastAPI, Streamlit, dbt, Feast feature store, FAISS, pgvector, model drift detection, Jupyter

Explainability & Responsible AI: SHAP, Explainable AI, fairness / bias auditing (Fairlearn)

Cloud: Google Cloud Platform, Vercel (foundational AWS exposure)

SELECTED PROJECTS

Agentic AI for Model Drift Detection in Medical Imaging | MSc Dissertation, University of Surrey | 2026 - Present

- Building an agent-based deep learning system that autonomously detects, diagnoses, and explains model drift in medical-imaging classifiers under distribution shift.
- Integrating Explainable AI attribution methods to surface drift root causes and trigger targeted retraining, sustaining classification accuracy as data distributions change.
- Designing monitoring workflows with automated alerting, model versioning, and retraining triggers for reliable production deployment.

Tools: *Python, PyTorch, Explainable AI, model monitoring, MLOps*

Vision Transformers for Video Action Recognition | Advanced Machine Learning, University of Surrey | 2026

- Built video action-classification pipelines with VideoMAE and TimeSformer using PyTorch and Hugging Face Transformers, achieving 90.4% top-1 accuracy on a 25-class HMDB benchmark.
- Ran controlled ablations (temporal sampling, augmentation, learning-rate schedules) and multi-seed robustness tests; spatial augmentation added ~1.6 percentage points of accuracy.

- Prototyped a DETR-style spatio-temporal action detector on a frozen backbone and analysed model behaviour with attention rollout and t-SNE visualisation.

Tools: *PyTorch, Hugging Face Transformers, VideoMAE, TimeSformer, TensorBoard*

CreditLens - Loan Default Risk & Decision Platform | Personal Project | 2026

- Developed an end-to-end credit-risk system with LightGBM and CatBoost against a Weight-of-Evidence scorecard baseline, with probability calibration, served through a FastAPI service.
- Re-framed the decision layer from accuracy-only to profit-optimal thresholds, converting an unprofitable lending policy into positive expected margin per applicant.
- Added SHAP explanations, adverse-action reason codes, Fairlearn bias audits, MLflow experiment tracking, a Feast feature store, and a 12-month simulated drift monitor.

Tools: *Python, LightGBM, CatBoost, scikit-learn, SHAP, Fairlearn, MLflow, Feast, FastAPI, DuckDB*

AskWarehouse - Text-to-SQL Analytics Agent | Personal Project | 2026

- Built a natural-language-to-SQL agent over a multi-table data warehouse using a locally hosted Qwen2.5-Coder LLM, DuckDB, and a dbt semantic layer.
- Implemented a self-correcting repair loop that feeds database errors back to the model, raising execution accuracy from 32% to 42% and valid-SQL rate to 87% on the BIRD benchmark.
- Added schema retrieval, value indexing, SQL AST validation, PII deny-lists, and ambiguity detection (95% accuracy) to keep generated queries safe and trustworthy.

Tools: *Python, Qwen2.5-Coder, DuckDB, dbt, sqlglot, Streamlit*

FilingsIQ - Verifiable Financial Question Answering (RAG) | Personal Project | 2026

- Built a retrieval-augmented QA system over SEC 10-K filings with hybrid BM25 + BGE-embedding retrieval and table-aware document parsing.
- Added arithmetic recomputation and provenance verification for derived metrics, cutting hallucination rate from 64% to 8% and reaching 92% accuracy on answered questions.
- Implemented refusal calibration so the system declines to answer rather than speculate when supporting evidence is absent.

Tools: *Python, sentence-transformers, BM25, SQLite, Qwen2.5, FastAPI*

Dialect-Robust Sentiment & Sarcasm Detection | NLP Project | 2026

- Fine-tuned BERT and RoBERTa transformers for sentiment and sarcasm classification across diverse English dialects, using targeted data augmentation to reduce dialect bias and evaluating fairness and generalisation.

Tools: *Python, PyTorch, Hugging Face Transformers, BERT, RoBERTa*

Additional projects: Eye Disease Detection - transfer-learning CNNs (ResNet50, VGG19, InceptionV3) in TensorFlow, ~95% test accuracy; OpenIncident - production Model Context Protocol server suite with OAuth2 gateway and 91 passing tests; Aegis - multi-agent AI platform with Neo4j knowledge graph and pgvector RAG; Medical RAG Bot - FastAPI + FAISS + LangChain assistant over clinical guidelines. Full code at github.com/dariyaDeepanshu.

EXPERIENCE

Artificial Intelligence & Machine Learning Intern | SmartInternz (Google Developers AI-ML Program), Remote | 2023

- Completed a Google Developers-backed applied AI/ML internship, building and evaluating supervised learning models in Python and scikit-learn on real-world datasets.
- Delivered an end-to-end project spanning data preprocessing, model training, hyperparameter tuning, and evaluation; earned external certification (credential Ext-AIML-2024-76652).

CERTIFICATIONS & CONTINUOUS LEARNING

- Artificial Intelligence & Machine Learning, Powered by Google Developers - SmartInternz (2023).
- LinkedIn Learning: 30+ completed courses in Deep Learning (Keras), Retrieval-Augmented Generation & LLMs, NLP, Advanced SQL for Data Science, and AI solution design.