Aditya Anil

Machine Learning Engineer

aditya8anil@gmail.com +61 451405868 Sydney, Australia https://skaarfacee.netlify.app/

Summary

Machine Learning Engineer experienced in building and deploying production AI systems, APIs, and computer-vision pipelines. At IHX, re-engineered a production product supporting approximately 90% of company revenue, achieving 99% table-classification precision and reducing processing time by 50%; in a personal project, built multimodal LangGraph agents with validation, tracing, and cost-aware model selection. Skilled in Python, PyTorch, AWS, FastAPI, and model evaluation, with ongoing research in interpretability at UTS.

Experience

IHX Bangalore, India

Machine Learning Engineer Jul 2023 – Jan 2025

IHX | Machine Learning Engineer

July 2023 – January 2025 | Bangalore, India

- Re-engineered a production document-intelligence product supporting approximately 90% of company revenue, improving processing efficiency, scalability, and operational reliability.
- Trained, fine-tuned, and integrated deep learning models including Donut, LayoutLMv3, Table Transformer (TATR), and DETR for healthcare document classification and structured information extraction.
- Built and optimised batch and real-time inference pipelines across CPU and GPU environments, balancing latency, throughput, and compute costs for different workloads.
- Achieved 99% precision in table classification and reduced pipeline execution time by 50% through parallel processing and workflow optimisation.
- Built Python and FastAPI REST APIs to serve AI-powered document-processing capabilities and integrate them with internal products and engineering teams.
- Supported production releases and deployments on AWS and GCP, administered servers via SSH, and monitored production systems with Grafana.
- Developed automated web-scraping and data-processing pipelines across four pharmacy platforms, creating a structured pharmaceutical data store and building a fraud-detection model using the collected data.
- Mentored and onboarded an ML engineer, documented production workflows, and transferred knowledge for critical systems.

UTS University of Technology Sydney (UTS) Sydney, Australia

Academic Tutor - 32130 Fundamentals of Data Analysis Sep 2025 – Apr 2026

- Evaluated assignments and led oral defenses in Fundamentals of Data Analysis to assess students' understanding.
- Applied probability, statistics, and data interpretation while maintaining academic integrity in the assessment process.
- Supported tutorials by explaining core statistical concepts and providing actionable feedback to strengthen students' analytical reasoning.

UTS University of Technology Sydney (UTS) Sydney, Australia

Research Assistant Feb 2025 – Present

- Mechanistic Interpretability of EEG Foundation Models: Investigating how pretrained EEG transformers (CBraMod) encode neurological states, using Cross-Layer Transcoders (CLTs), sparse feature representations, and activation-level analysis to identify interpretable neural features associated with drowsiness.
- Built end-to-end EEG machine-learning pipelines for preprocessing, model adaptation, feature extraction, and evaluation using NumPy, SciPy, and pandas; benchmarked linear probing, partial fine-tuning, LoRA, and full fine-tuning on the SEED-VIG dataset.
- Evaluated 153,600-dimensional representations across 12 transformer layers using hypothesis tests, effect sizes, permutation testing, and subject-independent validation to identify task-relevant features.
- Co-authored AE-UVNet, an attention-enhanced hybrid U-Net/V-Net architecture for multimodal brain tumour segmentation, published in Springer Lecture Notes in Computer Science (2026).
- Generative AI & Computer Vision Research: Investigated diffusion-based object detection for complex aerial imagery, focusing on robustness to low-resolution objects, occlusions, and challenging visual conditions.
- Co-authored three research publications spanning medical imaging, deep-learning model generation using matrix factorisation, and AI applications in clinical healthcare.

docTR Remote

Open Source Contributor Sep 2024 – Jan 2025

- Contributed to **docTR**, a leading open-source deep learning OCR framework (4.2K+ GitHub stars), improving text detection and recognition accuracy, model efficiency, and document understanding capabilities.
- Supported the optimisation of computer vision and deep learning pipelines for robust, high-performance optical character recognition.

Education

University of Technology Sydney (UTS) Masters • 84.33

Artificial Intelligence Feb 2025 – Oct 2026

Amrita Vishwa Vidyapeetham Bachelors of Technology • 87.00
Computer Science (Artificial Intelligence) Jun 2019 – Jun 2023

Projects

PitchSense - AI Football Analytics & Tactical Tracking Jun 2026

- Won Best Innovation/Research Project at the 2026 UTS AI Showcase
- Built an end-to-end computer vision system that converts standard football match footage into tactical telemetry without GPS trackers or specialised hardware, integrating YOLO-based models for player detection, ball detection, pitch keypoint estimation, and segmentation.
- Developed a video-tracking and pitch-registration pipeline using YOLO-Pose, ByteTrack, DLT homography, and smoothing to map player and ball positions onto a 105 m × 68 m tactical pitch; worked on tracking corrections to improve player-identity consistency during overlaps and camera movement.
- Built a Streamlit analytics interface with possession analysis, heatmaps, formations, territory control, passing networks, pressing analysis, player-movement statistics, and synchronized review videos.

Multimodal Memory & Agentic Content System Jun 2026

- Built Pixel Sprites, a Python multimodal AI application using LangChain and LangGraph, with specialised agents for task classification, media understanding, content planning, keyword generation, and output validation across images, video, and audio.
- Orchestrated stateful LangGraph workflows with conditional routing and retries; implemented structured Pydantic outputs and an agent-based validation step that checks generated content against requirements and retries failures up to two times.
- Configured OpenRouter model tiers for text, vision, and reasoning tasks to balance model capability and cost; integrated LangSmith tracing to inspect and debug agent execution, LLM calls, and workflow behaviour. Built media indexing and memory with SHA-256 identifiers, metadata extraction, structured JSON records, and Supabase persistence; used InsightFace for face detection and embedding clustering.

Habit Tracker & AI-Powered Task Categorizer Nov 2025

- Built and currently run an LLM-powered task-categorization agent in a full-stack habit-tracking system, integrating Google Tasks API to categorize tasks and assign difficulty levels.
- Designed a modular LLM provider architecture supporting Google Gemini and OpenAI APIs, enabling flexible model switching and scalable AI integration
- Built an android Flutter mobile application featuring an interactive productivity heatmap, dynamic theming (dark/light mode), and real-time task visualization
- Automated daily task categorization using GitHub Actions CI/CD workflows, securely managing API credentials via encrypted repository secrets.

AI Car Insurance Advisor Apr 2026

- Built an end-to-end AI insurance recommendation system combining driver-risk prediction, premium estimation, policy ranking, and evidence-backed recommendations.
- Developed predictive models using structured driver, vehicle, and policy features to estimate insurance risk and expected premiums.
- Implemented multi-criteria policy ranking using coverage compatibility, pricing, risk alignment, and insurer diversity to generate personalised recommendations.
- Integrated a Qdrant RAG pipeline to retrieve evidence from insurance Product Disclosure Statements and support explainable, evidence-backed recommendations.
- Developed a SQLite interaction store to capture user inputs, predictions, recommendations, and feedback, enabling analysis and future model improvements.

Skills

Agentic AI & LLM Engineering: LangGraph, LangChain, multi-agent systems, stateful agent orchestration, conditional routing, agent memory, LLM tool integration, structured outputs, Pydantic, agent-based validation, retries, OpenRouter, OpenAI API, Gemini API

AI Observability, Evaluation & Cost Optimisation: LangSmith tracing, LLM workflow debugging, model evaluation, output validation, cost-aware model routing, configurable LLM tiers, Grafana, latency optimisation, CPU/GPU workload scheduling

Production AI Engineering & MLOps: Python, FastAPI, REST APIs, deep learning model deployment, batch and real-time inference, production monitoring, Git, GitHub Actions, CI/CD, Linux, AWS, GCP

Deep Learning & Multimodal AI: PyTorch, TensorFlow, Hugging Face Transformers, fine-tuning, LoRA, vision-language models, Donut, LayoutLMv3, DETR, Table Transformer, YOLO, OpenCV, Whisper, InsightFace, OCR, document intelligence

RAG, Data & Persistence: Retrieval-Augmented Generation (RAG), Qdrant, Supabase, Firebase, SQLite, SQL, NumPy, pandas, SciPy, vector embeddings, metadata extraction, structured JSON

Applications & Research: Streamlit, Flutter, scikit-learn, statistical analysis, mechanistic interpretability, sparse feature analysis, model benchmarking