

Abdallah Muhammed

Mobile: 07827241552 Email: abdallahmuhammed001@gmail.com

LinkedIn: www.linkedin.com/in/abdallahmuhammed GitHub: www.github.com/daddal001

EDUCATION AND QUALIFICATIONS

University of Sheffield **2022 – Present**

MEng Mechatronics and Robotic Engineering with an Industrial Placement Year

- **Achievements:** Predicted First Class Honours
- **Course Includes:** Machine Learning, Systems Modelling, State Space Analysis, Control Techniques, Engineering Mathematics, Signals Processing, Computer Systems Design

The National Mathematics and Science College and Oxford International College

A Levels: Mathematics (A*), Physics (A), Computer Science (B), Further Mathematics (B) **2022**

IGCSEs: 8 (A*-A) including Mathematics (A*), English Language (A), English Literature (A) **2019**

EXPERIENCE

IKOS AEGIS, Innovation Engineer **2025 – Present**

- Sole engineer of a production document intelligence agentic AI platform; architected, built, and operated end-to-end: a Next.js/React frontend, ten-plus Python (FastAPI) and Go microservices, Vertex AI/Gemini inference, and full IaC and observability (Terraform, ArgoCD, OpenTelemetry, Grafana) on GKE Autopilot.
- Utilised an AI-augmented development workflow, in tandem with my open-source project the Universal AI-driven documentation standard to create proper documentation, allowing for shipment of production code faster without accumulating technical debt.
- Built a high-throughput document pipeline, with resumable batch uploads, ClamAV malware scanning with quarantine/promotion, and two-stage parsing (two-tier parser). Hardened with queue backpressure, cooperative preemption, and idempotent, refresh-resilient task recovery. Processes 1,000+ documents per batch and cuts manual extraction by up to four days.
- Built the platform's LLM layer on Vertex AI Gemini (langchain-google-vertexai) using schema-constrained structured generation backed by a Pydantic schema registry that serves as the model's generation-time wire contract, with streaming responses, multimodal text and image inputs, and contract tests guarding against malformed-output failure modes.
- Designed a confidence-gated, multi-tier extraction cascade that escalates context (partial text → + images → full text → + full images) only when model confidence is low, with best-of selection and human-in-the-loop review for low-confidence outputs, bounding token cost while lifting accuracy on complex documents.
- Engineered production-grade LLM reliability using jittered exponential backoff with rate-limit-aware circuit breaking, a per-pod inference-concurrency limiter exposed through Prometheus saturation metrics, model-output validation, and fail-closed service boundaries.
- Implemented zero-trust security and compliance-aligned governance, using Workload Identity, default-deny network policies, a BFF auth layer (Keycloak OAuth 2.1 + PKCE with OWASP fingerprint binding), HMAC request signing, and a CI security gate (Semgrep, Trivy, TruffleHog, Bandit) mapped to SOC 2 / ISO 27001 / EU AI Act controls.
- Operated the platform through real production incidents such as schema-contract failure, event-loop starvation, storage race conditions, driving the fixes through documented architecture decisions and post-mortems.
- Delivered an engineering-assessment tool projected to save £20,000+ annually by replacing outsourced assessment workflows.

Moniepoint Microfinance Bank Limited, Data Engineer (Part-Time) **2023 – 2025**

- Implemented machine learning models for business classification, achieving a final accuracy of ~90%, resulting in improved targeted marketing strategies, leading to a measurable increase in customer satisfaction.
- Reduced database query time by 6x through optimised data pipelines, enhancing operational efficiency.
- Leveraged Google Cloud virtual machines to accelerate algorithm training, reducing computation time by 5x.
- Introduced a large language model (LLM) approach to select businesses for an in-house newsletter, improving relevance and user engagement over the previous rules-based approach.
- Automated employee ID generation by implementing new data pipelines and dashboards, increasing process efficiency and eliminating manual ID-generation errors.

University of Sheffield, Team Lead (Predictive Maintenance Algorithm Project) **2024 – 2025**

- Spearheaded a team of 9 to develop a predictive maintenance solution using Digital Twin (DT) and Industrial IoT (IIoT) technologies for real-time monitoring and anomaly detection in cyber-physical systems.
- Built and tested a 3-layered machine learning pipeline using RNN-based variational autoencoders, clustering

and temporal trend analysis to analyse sensor data, achieving fault detection and predictive accuracy without pre-existing failure data.

- Designed a scalable, cloud-integrated system that complies with Industry 4.0 standards, focusing on reducing downtime and improving system reliability.
- Secured 2nd Place in Siemens Connected Curriculum Competition, recognised for technical depth and Industry 4.0 alignment.

OPEN-SOURCE PROJECTS

Two-Tier Document Parser (deployed in production within the IKOS AEGIS platform)

- Built a production document-parsing service that routes between a fast CPU tier (PyMuPDF4LLM, sub-second per page) and a high-accuracy MinerU VLM tier (~95% on complex layouts, tables, and formulas), with automatic GPU-to-CPU fallback for hardware-agnostic deployment.
- Engineered it for sustained concurrent load with isolated process pools and queue-depth backpressure that ring-fence the latency-sensitive preview route from bulk parsing, cooperative request preemption that resumes from the last completed page on cancellation, and periodic GC/CUDA-cache reclamation that holds the long-lived in-process VLM's memory steady across thousands of parses.
- Hardened the service with OWASP ASVS upload limits and SHA-256 image deduplication, instrumented it with Prometheus and OpenTelemetry and containerized it for deployment on GKE Autopilot.

Universal AI-Driven Documentation Standard (Documentation Standard for AI Coding Agents)

- Built an open-source framework that grounds AI coding agents in a project's own documentation to cut hallucination and regressions.
- Scope-aware hooks and generated agent rules (.cursorrules, CLAUDE.md, Codex instructions) force agents to keep docs current as they code, then have them verify against those docs when planning instead of guessing at APIs, configs, or prior decisions.
- Wired it into CI with shell validators for frontmatter, structure, link integrity, staleness, PII, and readability, plus a one-command tiered bootstrap (solo → enterprise). Used in the IKOS AEGIS platform to ship production code faster without the accumulating technical debt.

TECHNICAL WRITING

Selected engineering essay published on LinkedIn, on production AI reliability and platform patterns:

- Your Files Vanish When You Refresh: A TOCTOU Race Between MinIO and Redis - diagnoses a cross-store consistency bug where a read endpoint deleted Redis inflight state while object-store listing lag made uploaded documents temporarily disappear.

EXTRA-CURRICULAR ACTIVITIES

Imaging Systems Engineer – Project Hex (SUAS Competition)

2023 – 2024

- Developed computer-vision algorithms for an autonomous drone in the Student Unmanned Aerial Systems (SUAS) competition, achieving 90% accuracy in object recognition.
- Designed advanced algorithm pipelines integrating image preprocessing (denoising, scaling), feature extraction (HOG, SIFT), and classification using Convolutional Neural Networks (CNNs).
- Implemented techniques such as transfer learning and model fine-tuning to improve performance on domain-specific data.
- Optimised real-time mapping pipelines for drone navigation, incorporating Kalman filters for position estimation of ground coordinates from the video feed.

SKILLS AND INTERESTS

Languages: Python, SQL, C/C++, MATLAB

Generative & Agentic AI: LLM application engineering, agentic workflows, RAG, tool use, model-output evaluation, prompt engineering, LangChain, LangGraph

Machine Learning & Classical AI: Deep learning (CNNs, RNNs/GRUs, VAEs, transformers), computer vision (OpenCV, Tesseract OCR), anomaly detection, time-series analysis, dimensionality reduction (PCA, t-SNE), clustering, PyTorch, TensorFlow, scikit-learn, pandas, NumPy

Backend & MLOps: FastAPI, Docker, Kubernetes, Redis, PostgreSQL / TimescaleDB, S3-compatible object storage, task queues, REST APIs, system architecture design

Cloud, DevOps & Observability: Google Cloud Platform, Vertex AI, GKE, Terraform, Kustomize, ArgoCD, Git, OpenTelemetry, Prometheus, Grafana, Loki, Tempo, audit logging

Societies: ACSE Society, Ethical Hacking Society, Finance Society

Hobbies: stock market trading, crypto trading, hiking, geopolitics, polo