Noel Alex Samuel

AI / ML ENGINEERING · DATA SCIENCE · DATA SYSTEMS

noelalex404@gmail.com LinkedIn github.com/Noel-Alex Portfolio: noel-alex.github.io/portfolio

PROFILE

AI engineer and final-year B.Tech CSE (Data Science) student with six months of professional experience building GPU-served document AI. Leads R&D across graph ML, trustworthy evaluation and scalable data pipelines, from experiment design to deployment, performance optimization and mentoring.

PROFESSIONAL EXPERIENCE

AI Engineer (contract) — Blackridge Technologies / Stravah Mar–Sep 2026

- Built AI workflows that extract and reconcile transactions from complex Indian bank-statement PDFs: layout and table reconstruction, OCR, row recovery and ledger consistency checks.
- Served GLM-OCR on GPUs with FastAPI, vLLM and Modal (L40S), tuning batching, concurrency, caching, observability and cost, backed by load and quality tests.

SELECTED PROJECTS & RESEARCH

Stretchy + HALO — Evidence-grounded AI platform and embedding research 2025–26

- Led R&D on a fault-tolerant Python/Rust pipeline spanning document extraction, PostgreSQL orchestration, Qdrant retrieval and provenance-aware processing.
- Inventoried 862,917 files (130.26 GB) with zero scan errors via resumable bounded-memory scanning (~147 MiB near 747K files); HALO cut encoder-token input ~5.1× on 2,536 real edits versus a late-chunking baseline.

Arachne — Configurable, high-performance web crawler in Rust 2025–26

- Built site, sitemap, list and forensic-capture crawling on NATS JetStream with SQLite and content-addressed storage, headless-Chrome rendering, SSRF-checked redirects and volunteer workers over Tailscale; collected 181,755 live captures.

PhishGCN — Phishing detection and label-quality research 2026

- Found 64.7% of phishing-reported live captures were stand-ins, not phishing pages; graph models trained with label-quality controls reached 0.724 PR-AUC (0.847 ROC-AUC) on 1,196 human-audited captures versus 0.614 on feed labels, under site/kit-disjoint evaluation.

PolySafe / Vergil — Reliable drug-interaction ML 2026

- Led GNN/Chemprop experiments on 63,472 drug pairs and 964 adverse-event classes; designed unseen-drug and scaffold-OOD splits and tested calibration, weak generalization and statistical abstention limits.

VCAP / Living Review — Faculty-gated AI code review 2026

- Co-developed versioned feedback, provenance and human-approval workflows; co-authoring research in preparation on code-comment relocation across 7,200 controlled edits and 2,073 real commit pairs.

HELIX Lite — Parallelizing compiler in Rust 2026

- Wrote the full pipeline (Pratt parser, SSA, dependence tests, Cranelift JIT) to prove loops safe before parallelizing them: 5.79× saxpy speedup on 8 threads, with unsafe loops refused and explained.

Sivana — Audio recognition and indexed matching 2025–26

- Built a fingerprint matching engine and browser-local WASM flow; calibrated to 76.7% gated recall versus 35% legacy recall at zero false accepts on a small degradation benchmark.

EDUCATION

Vellore Institute of Technology, Vellore B.Tech, Computer Science & Engineering (Data Science) · 2023–27

New Middle East International School, Riyadh CBSE (Classes X and XII) · Physics subject topper · 2023

LEADERSHIP & ACHIEVEMENTS

GDG VIT · Senior Technical Member, AI/ML (2023–26) Helped run DevJams and Women Techies, ran AI/ML recruitment end to end (256 shortlisted from about 5,000 applicants, about 80 given a project and interviewed by me), and mentored juniors.

Inside Google's DeepMind (2024) Designed and gave a recorded session for 400+ students, from transformers to AlphaFold.

Head Boy · New Middle East International School (2022–23) Coordinated student welfare, staff liaison and off-campus safety during the return to in-person school, and represented the school at a Middle East CBSE principals' conference.

TECHNICAL SKILLS & CERTIFICATION

Languages Python, Rust, SQL, TypeScript, C/C++, Java, R

ML PyTorch, Transformers, GNNs, NLP / OCR, RAG, evaluation design, pandas, scikit-learn

Systems FastAPI, vLLM, Modal, PostgreSQL, NATS JetStream, Qdrant, Docker, Temporal, Linux

Certification Oracle Cloud Infrastructure 2025 Generative AI Professional (92%)