

# Seif Ali

Chippendale NSW 2008, Sydney · +61 422 337 484 · seifarafa@hotmail.com · linkedin.com/in/seif-arafa · github.com/sa1701

## SUMMARY

---

Final-year Computer Science student at the University of Wollongong (Artificial Intelligence & Big Data major), graduating December 2026. Building production-grade AI systems end-to-end — local-first RAG pipelines with multimodal LLMs, full-stack web applications, from-scratch ML classifiers, and CI-tested PHP/MySQL backends. Currently completing a year-long capstone in software-update security analysis, combining program analysis with transformer-based NLP to detect silently-patched vulnerabilities.

## EDUCATION

---

### Bachelor of Computer Science — Artificial Intelligence & Big Data

University of Wollongong, Australia · September 2023 – December 2026 (expected)

- Commenced at the University of Wollongong's Dubai campus (2023); transferred to the main Wollongong campus, Sydney (2025).
- Relevant coursework: Big Data Mining; Software Development Methodologies; Generative AI; Modern AI; Foundations of AI; Algorithms & Data Structures; Database Systems; System Analysis & Design; Cyber Security; Human–Computer Interaction; Astrophysics; Final-Year Capstone Project.

## PROJECTS

---

### GEARHEAD — Local-First AI Diagnostic Assistant

Python · FastAPI · Multimodal LLM · ChromaDB · PyMuPDF · Retrieval-Augmented Generation

- Built an end-to-end diagnostic assistant grounded in real workshop manuals: a user describes a symptom or enters a fault code, and the system returns a cited fix alongside the relevant manual diagram.
- Ingested 22 service manuals into a local Chroma vector store (2,450 chunks) with page-level citations on every answer.
- Designed a swappable AI-provider interface so the underlying language model can be substituted without code changes (commercial APIs or locally-hosted open-weight models via Ollama); renders manual pages as images for multimodal vision interpretation.
- Implemented prompt caching across the system context, cutting per-query token cost by approximately 90%. Delivered a FastAPI web app validated by an end-to-end test suite.

### ZynTrace — Software Update Tracing System (Final-Year Capstone, in progress)

Python · Natural Language Processing · Program Analysis · Transformer Models

- Developing an automated tool that diffs code changes against release notes to detect bugs, silently-patched security vulnerabilities, and reused code patterns (clones) that propagate the same defect across a codebase.
- Combines static program analysis on commit diffs with transformer-based NLP to align textual summaries with code-level changes; designed for evaluation against CVE-fixing corpora (BigVul, CVEFixes).

### RAG PDF Chat — Privacy-Preserving Document Q&A

Python · LangChain · FAISS · sentence-transformers · Ollama · Streamlit

- Built a privacy-first document analysis tool: users upload a PDF and receive cited answers from a locally-hosted language model, with no data leaving the machine — suitable for sensitive or confidential documents.
- Implemented the full retrieval pipeline: PDF extraction, semantic chunking, vector embeddings (all-MiniLM-L6-v2), FAISS cosine-similarity retrieval, and grounded generation via Ollama.

### Talent Matching Platform — Group Software Engineering Project

PHP 8 · MySQL 8 · Object-Oriented Design · PHPUnit · GitHub Actions CI

- Backend developer in a five-person team building a job-candidate matching platform; owned authentication, job postings, recommendation engine, and membership tiers across 4 of 14 release commits.
- Designed a weighted scoring engine (55-point scale) and fuzzy keyword search using Levenshtein distance and Soundex; configured a CI pipeline running lint and PHPUnit on every push.

### TaskFlow — Full-Stack Kanban Application

Next.js · TypeScript · Supabase · Prisma · NextAuth.js · Tailwind CSS

- Designed and deployed a Kanban task manager with drag-and-drop, multi-project support, and JWT authentication; deployed in production on Vercel with Supabase-hosted PostgreSQL.

## **Big Data Mining — Classifiers & Regression from First Principles**

*Python · NumPy · pandas · scikit-learn · Spark MLlib · TensorFlow*

- Implemented two decision-tree classifiers from scratch (Information Gain and Gini Index, no ML libraries) on the UCI Nursery dataset (12,960 instances), with weighted ensembling for class imbalance.
- Built an end-to-end regression pipeline on UCI Online News Popularity (39,644 articles, 58 features) using log-transformation and Spearman rank correlation. Coursework covered Spark/MapReduce, naive Bayes, ANNs, CNNs, PageRank, and streaming data mining.

## **CS Clash — Terminal Strategy Game in C (in progress)**

*C · Procedural Programming · Test-Driven I/O Validation*

- Building a Clash of Clans-inspired strategy game in pure C (standard library only): a Barbarian raids a 10×10 village defended by cannons, archer towers, bombs, and walls to destroy the Town Hall.
- Implemented the command-driven setup engine — entity placement with bounds and occupancy validation, multi-tile wall runs, and single-Town-Hall enforcement with prioritised error handling — verified against a suite of stdin/stdout test cases.

## **TECHNICAL SKILLS**

---

- **AI / ML:** RAG, LangChain, FAISS, sentence-transformers, Hugging Face Transformers, multimodal LLM integration (commercial APIs and locally-hosted via Ollama), prompt caching, scikit-learn, TensorFlow, Spark MLlib, CNNs, RNNs, NLP, prompt engineering.
- **Languages:** Python, TypeScript, JavaScript, C, C++, Java, PHP, SQL, HTML, CSS.
- **Backend:** FastAPI, Node.js, Express, Next.js, Prisma, REST API design.
- **Frontend:** React, Next.js, Tailwind CSS, Zustand, Framer Motion.
- **Databases:** PostgreSQL, MySQL, MongoDB, Supabase, ChromaDB.
- **Data & DevOps:** NumPy, pandas, Spark / MapReduce, PyMuPDF, Jupyter, Git, GitHub Actions, Docker, PHPUnit, Streamlit, Vercel.

## **LANGUAGES**

---

<b>Arabic</b>	Native
<b>English</b>	Fluent (IELTS overall band 7.5)