Ehis Oghina

ehisoghina.com | LinkedIn | GitHub

Education

University of Birmingham — Birmingham

Master of Science in Artificial Intelligence — *Sep 2026 – Sep 2027*

Dublin City University — Dublin

Bachelor of Science (Honours) in Computer Science — *Aug 2022 – May 2026*

- Key Modules: Artificial Intelligence, Advanced Algorithms, Probability & Statistics, Linear Algebra

Experience

NLP Researcher — Sep 2026 – Present

NLP Research Group, University of Birmingham — *Birmingham*

- Selected to join a university research team participating in SemEval 2027, investigating cross-lingual alignment and activation steering techniques for large language models.
- Contributing to collaborative NLP research focused on multilingual language generation, model interpretability and evaluation of language-steering methods.

Software Engineer / Data Scientist — Apr 2025 – Oct 2025

Optum, UnitedHealthCare — *Dublin*

- Built a Snowflake-backed Dash analytics platform for real-time exploration of application performance and analytics trends. Reduced manual reporting effort by seven hours per week (70%) and supported Team Leaders and Data Analysts with custom filtering, query execution, and visual dashboards.
- Implemented RapidFuzz-based string matching across large pharmaceutical datasets to flag inconsistent, duplicate, or mismatched records, improving data quality by 56%. Led stakeholder meetings across business and engineering teams, translating requirements into Databricks and Snowflake deliverables, including waste analysis reports.

R.V. Uncovered – Founder & Lead Engineer — June 2026 – Present

Independent Venture — *Birmingham*

- Founded a football analytics platform that adjusts player and team metrics for match context, including game state, red cards, substitutions and injuries. Exposes misleading headline statistics such as raw xG and distance covered.
- Designing a tool-calling LLM agent that answers natural-language football questions with grounded data, visualisations and caveats. Developing a computer-vision pipeline using pitch calibration, player tracking and shirt-number identification to generate tracking data from amateur match footage.

Software Engineering Associate — Jan 2025 – May 2025

EPAM Systems — *Dublin*

- Prototyped Raspberry Pi-based IoT automation systems in Python, implementing GPIO sensor control and cloud API synchronisation to support real-time device monitoring and automated data capture.
- Built lightweight data pipelines on embedded Linux using Pandas, Matplotlib, and Dash to collect, process, and visualise sensor data. Supported testing, documentation, and prototype refinement in agile workflows to improve reliability and maintainability.

DCU Soccer – IT Specialist, Club Chairperson — Aug 2022 – Jun 2026

Dublin City University — *Dublin*

- Managed day-to-day operations for a university football club with 500+ members, overseeing budgeting, fixture coordination, event planning, and communication with DCU Athletics and external clubs.
- Managed club IT infrastructure and internal communication systems, improving coordination and reducing reliance on manual messaging and fragmented administration.

Projects

Little by Little - Grocery Optimisation

Oct. 2026 – Present

Python, FastAPI, PostgreSQL, OR-Tools, TypeScript, React, LLM API, Maps

- Building a full-stack web app that helps Irish households decide where and how to do their weekly grocery shop, comparing single-store and split-shop plans across major supermarkets while accounting for loyalty vouchers, travel cost and time. The Cheapest, Fastest and most Balanced options are returned with a store-by-store checklist.
- Designing a deterministic optimisation engine for pricing, voucher eligibility and product matching, using an LLM only to parse shopping lists and explain recommendations in plain language.

Trivium – Tournament Analytics Platform

Sep. 2026 – Nov. 2026

Python, Pandas, DuckDB, Scikit-learn, Statsmodels, Streamlit, Plotly, Claude API

- Built a football analytics pipeline on match event and tracking data from the FIFA U-17 World Cups 2025, men's and women's. Defined three team scores (dominance, efficiency and resourcefulness), standardised within each tournament and measured against a pre-tournament Elo expectation, to identify which tactical profiles take teams further than their strength predicts.
- Delivered a Streamlit dashboard with team reports, match reviews and a coach scouting view, backed by a natural-language query box that sends only aggregated results to the Claude API and falls back to offline RapidFuzz matching. Validated findings with out-of-sample, robustness and sensitivity tests, and reported team-level results only.

Social Olympics – Competition Platform

Jun 2026 – Oct 2026

TypeScript, Next.js, React Native, Expo, Supabase, PostgreSQL

- Built a cross-platform web and mobile app for running custom Olympic-style competitions. Shared types, validation, scoring logic, and UI packages between Next.js and React Native (Expo) clients using a Turborepo monorepo on a Supabase backend.
- Designed a configurable scoring engine across a 52-event library in 16 categories. Combined tiered placement points, event weight multipliers ($0.5\times$–$3\times$), best-of-N selection, MVP bonuses, multi-stage tiebreakers, and snake-draft team balancing from peer-voted strength ratings.

xgGaffer – Sports Analytics Platform

Aug 2025 – May 2026

Python, Django, React, TypeScript, PostgreSQL, Scikit-learn

- Built a football analytics platform across five European leagues on an ML pipeline that scraped 43,000+ matches and 64,000+ player records. Engineered 86+ features and trained seven models covering match outcomes, scorelines, season simulation, injury risk, tactics, and lineups.
- Engineered a deterministic five-layer natural language query pipeline classifying 16 intent types via TF-IDF and Logistic Regression, routing to the ML engines. Used the Claude Haiku API only as a grounded formatter to prevent hallucination. Delivered role-based tools for coaches, referees, players, and fans, including a referee education module built by parsing the FIFA Laws of the Game with Gemini 2.5 Flash Lite.

Financial Risk Dashboard

Nov 2024 – Feb 2025

Python, Django, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

- Developed a Django application that cleans historical market data with Pandas and NumPy and computes portfolio-level risk metrics. Presented risk trends and portfolio performance through interactive Matplotlib and Seaborn visualisations.
- Trained Scikit-learn models to predict Value at Risk (VaR) and volatility patterns, achieving 82% accuracy across a 5-year historical dataset.

Technical Skills

ML & Data Science	Scikit-learn, Pandas, NumPy, Matplotlib, Seaborn, LLMs (AWS Bedrock, OpenAI API), Data Cleaning, Fuzzy Matching (RapidFuzz)
Languages	Python, Java, C/C++, JavaScript, TypeScript, SQL, R, HTML/CSS
Frameworks & Cloud	Django, React, Flask, Dash, Docker, AWS (Lambda, Bedrock), Snowflake, PostgreSQL, Databricks, GCP
Tools & Testing	Git, GitHub, Postman, Pytest, Selenium, Jira, Linux