

MSc candidate in Artificial Intelligence in Biomedicine and Healthcare at UCL, passionate about applying my expertise to complex biomedical and healthcare challenges. I'm a self-motivated, adaptable learner who excels at collaborating with cross-functional teams and rapidly mastering new AI tools and frameworks.

EDUCATION

MSc Artificial Intelligence in Biomedicine and Healthcare , <i>University College London(UCL)</i>	09 2024 — 09 2025
BSc Physics , <i>University College London(UCL)</i>	09 2021 — 09 2024

SKILLS

Programming	Python, SQL.
AI & ML	PyTorch, scikit-learn, XGBoost, CatBoost.
NLP & LLMs	HuggingFace Transformers, Retrieval-Augmented Generation (RAG), Entity linking.
Vision & Imaging	Torchvision, CNN architectures, Variational Autoencoders (VAEs).

SELECTED PROJECTS

LLM-Based Entity Linking (MSc Dissertation) — [Most details](#) 06 2025 — Present

- Building an LLM-based entity linking system to map mentions in biomedical documents to knowledge base concepts. The model will jointly perform mention detection, fine-grained entity typing, and entity disambiguation in a single forward pass.
- Performed mention detection via BIO tagging over BioBERT token embeddings. Then implemented fine-grained entity typing with a multi-label classifier on detected mention representations.
- Will integrate FAISS to index and search a large-scale biomedical knowledge base for rapid candidate generation per mention.
- Will fuse entity-type similarity and description similarity via a linear layer to compute a scalar score for each mention–entity pair, then select the highest-scoring candidate.

Legal Q&A Assistant LLM Chatbot (B.E.R.T) — [View on GitHub](#) 01 2025 — 05 2025

- Developed an end-to-end Retrieval-Augmented Generation (RAG) pipeline for legal QA, including query translation, optimized retrieval system and GPT-4–based response generation.
- Matched or exceeded benchmark model(LegalBenchRAG) using open-source embeddings (SBERT, GTE-Large), RCTS chunking, and cosine-similarity search—achieving top Precision@K & Recall@K without proprietary APIs.
- Adapted retrieval depth and answer style via query complexity (vague vs verbose) and user expertise (layperson vs expert) classifiers, significantly reducing hallucinations and tailoring legal explanations.
- Validated with both IR metrics (Precision@K, Recall@K) and generation metrics (BERTScore-F1, faithfulness), demonstrating a scalable, reproducible pipeline for accurate legal advice.

Chest X-Ray Pneumonia Classifier — [View on GitHub](#) 11 2024 — 12 2024

- Preprocessed Chest X-Ray Pneumonia dataset (5.8 K images) and created custom DataLoaders with resizing, normalization, and balanced train/val/test splits.
- Implemented a baseline CNN (2 conv + ReLU + pooling) that achieved 89.6 % test accuracy.
- Added real-time data augmentation (flips, rotations, color jitter), batch normalization, and dropout to enhance model robustness.
- Introduced early stopping to prevent overfitting and boosted accuracy to 93.9 % (Normal: 83.0 %, Pneumonia: 97.9 %) while reducing test loss from 0.2588 to 0.1850.

Patient Readmission Prediction — [View on GitHub](#) 05 2025 — 06 2025

- Preprocessed 101K records with 50 features: dropped high-missingness fields, binarised rare classes, applied one-hot/ordinal/target encoding with -fold to avoid leakage, and standardised for SVM.
- Built Linear SVM, Random Forest and XGBoost models under stratified 5-fold CV; tuned hyperparameters via GridSearchCV & RandomizedSearchCV.
- Random Forest Classifier performed best (CV F1 = 0.678, test F1 = 0.590), with key predictors: number of times patient was admitted and diagnostic codes.
- Reduced to top 25 features via importance-based selection, preserving performance while improving interpretability.

CERTIFICATIONS

IBM AI Engineering Professional Certificate (Coursera)	04 2025 — Present
--	-------------------