

Zhiwen Chen

+44 7756 574521 | zhiwen.c7777@gmail.com | linkedin.com/in/zhiwen-chen | zhiwen-chen-uk.github.io | London, UK

PROFESSIONAL SUMMARY

MSc Health Data Science student at UCL with a 2:1 BSc Computer Science from the University of Birmingham, targeting early-career AI, NLP, data-analysis, and healthcare-data roles in the UK. Hands-on experience across machine learning, natural language processing, OCR pipelines, and multi-objective optimisation, with a research artefact (BSc dissertation on neural architecture search) and an industry NLP internship behind it. Comfortable turning research papers into working PyTorch code, and equally at home in a transformer notebook or a clinical ontology. Looking for an internship, graduate, or junior role where I can apply ML and NLP to real problems, especially in healthcare, digital health, and applied AI.

CORE COMPETENCIES

Machine Learning	Natural Language Processing	Information Extraction	Clinical NLP / SNOMED CT	
Multi-objective Optimisation	Neural Architecture Search	Statistical Evaluation	Python & PyTorch	GWAS / Bioinformatics

WORK EXPERIENCE

Shanghai Zhiyu Information Technology Co., Ltd.

Jun 2025 - Sep 2025

NLP Algorithm Intern - Beijing Branch, China

- Built OCR and NLP pipelines for automated identification of lending-related clauses in scanned corporate charter documents, including text extraction, region localisation, and document cleaning.
- Developed and refined model components for clause classification and information extraction over noisy scanned PDFs.
- Iterated on data processing and experimental tuning to improve document analysis accuracy and end-to-end pipeline efficiency.

Allink Co., Ltd.

2020 - 2021

Frontend Developer Intern - China

- Developed frontend features for WeChat Mini Programs and collaborated with backend developers on API integration and error handling.
- Used Git and GitHub for version control and team collaboration on production codebases.

SELECTED PROJECTS

SENAS: Cyclical Three-Phase Evolutionary Algorithm for Multi-Objective Neural Architecture Search BSc Dissertation

University of Birmingham, Sep 2024 - May 2025 · Supervisor: Dr. Jizheng Wan · Python, PyTorch, EvoXBench

- Designed **SENAS**, a cyclic three-phase evolutionary NAS algorithm alternating between SOM-guided diversity exploration, dual-archive (CA/DA) convergence-diversity co-evolution, and ϵ -indicator-guided refinement.
- Introduced a lightweight Monte-Carlo sampling estimator for hypervolume, enabling dynamic early stopping and phase transitions without computing exact HV.
- Outperformed NSGA-II, MOEA/D, HypE, RVEA, and a SOM-SEMO baseline on **6 of 9** C10/MOP benchmark tasks (objective dimensionality 2D-9D) across NAS-Bench-101/201, NATS-Bench, and DARTS; e.g. HV 0.9368 on C10/MOP1 vs NSGA-II 0.9156 and HypE 0.8060 (30 independent runs, 10,000 architecture-query budget).
- Implemented non-dominated sorting, SOM-based neighbourhood exploration, dual-archive co-evolution, and a sigmoid-controlled crossover operator from scratch in Python with PyTorch.

Ontology-aware Clinical Entity Linking with SNOMED CT UCL MSc Research

University College London, ongoing · Python, Transformers, SNOMED CT

- Developing a two-stage entity linking framework for clinical text covering candidate generation and concept disambiguation against the SNOMED CT ontology.
- Exploring ontology-aware reranking strategies that combine semantic similarity with hierarchical signals such as hierarchy distance and shared-ancestor features.
- Designing experiments to assess hierarchy distance, shared-ancestor signals, and structural consistency in disambiguation performance.

Heart Failure Prediction from Clinical Features UCL MSc Coursework

University College London, 2026 · Python, scikit-learn

- Built an end-to-end machine learning workflow to predict heart failure status from structured clinical variables.
- Performed exploratory data analysis, missingness review, and comparative feature visualisation to support model design.
- Evaluated classification models using ROC-AUC, precision, recall, and F1-score, with emphasis on clinically relevant feature interpretation.

EDUCATION

MSc Health Data Science - University College London

Sep 2025 - Sep 2026 (expected)

Relevant modules: Artificial Intelligence in Healthcare, Advanced Methods in Data Science and Statistics, Principles of Health Data Science, Applied Computational Genomics, Epidemiology for Data Science, Programming with Python for Health Research.

BSc Computer Science, 2:1 - University of Birmingham

Sep 2021 - Jun 2025

Selected modules: Artificial Intelligence, Natural Language Processing, Evolutionary Computation, Human-Computer Interaction, Data Structures and Algorithms, Operating Systems, Software Engineering, Full-Stack Application Development, Functional Programming, Mathematics and Logic Foundations.

Educated at UK universities (BSc and MSc) with English as the language of instruction.

TECHNICAL SKILLS

Programming: Python, Java, JavaScript, C, Haskell, R

Machine Learning & NLP: PyTorch, Transformers (Hugging Face), PaddleNLP, information extraction, model fine-tuning, evolutionary algorithms, multi-objective optimisation, neural architecture search

Data & Research: experimental design, benchmarking, optimisation, error analysis, statistical evaluation, exploratory data analysis

Genomics & Bioinformatics: PLINK, GWAS analysis, variant QC, association testing

Tools: Git, GitHub, Linux/Unix, OCR tools