Amelia Ghanea

in linkedin.com/in/amelia-g   github.com/Amelia3141   ✉ amelia.ghaneahercock.24@ucl.ac.uk

📞 07341384667   🔗 amelia3141.github.io

🪪 PROFILE

- First Year MPhil/PhD student at UCL Institute of Health Informatics developing machine learning frameworks for rare disease research.
- First Class Neuroscience BSc Hons from the University of Bristol.
- Recipient of the 2024 Snowdon Master's Scholarship and member of the Windsor Fellowship.

🎓 EDUCATION

Health Informatics MPhil/PhD,   05/2025 – Present
Institute of Health Informatics, University College London

- Research focus: Machine learning frameworks for understanding complex disease relationships in the EDS-POTS-MCAS triad
- Current projects: Large-scale web scraping of patient self-report data from StuffThatWorks.com 🔗 (2.9M members, 110M data points across 1,263 conditions) to create temporal datasets for rare disease ML analysis
- Developing multi-modal GARNN architecture integrating heterogeneous features (symptoms, treatments, demographics) for disease trajectory prediction
- Implementing interpretable GARNN variants with attention visualization to identify distinct patient subgroups within the disease triad
- Exploring causal attention mechanisms in GARNN to distinguish correlation from causation in comorbidity patterns

Neuroscience BSc: First Class Honours with PLUS Award,   2021 – 2024
University of Bristol

💼 EXPERIENCE

ML Research Scientist, *ADVAI*   04/2025 – Present

- Led novel research on bias detection in medical LLMs using mechanistic interpretability techniques with Gemma 2B IT model and Sparse Autoencoders (SAEs)
- Developed innovative "feature clamping" methodology to replicate prompt-level biases through internal model interventions, multiplying specific SAE feature activations by factors (α=5.0-10.0) to simulate demographic presence without prompt modifications Implemented experimental pipeline using TransformerLens hooks on layer 12 residual streams (blocks.12.hook_resid_post) to extract and manipulate 16,384-dimensional SAE features in realtime
- Designed three-condition experimental framework comparing: (1) demographic-inclusive prompts, (2) demographic-scrubbed prompts, and (3) scrubbed prompts with internally clamped features on DDXPlus medical dataset

- Achieved measurable bias replication through internal interventions, with clamped features producing diagnosis shifts matching or exceeding prompt-based demographic effects (Jaccard similarity >0.7 between activation patterns)
- Evaluated bias using demographic parity and equalized odds metrics, demonstrating that internal feature manipulation can amplify demographic biases beyond prompt-level effects
- Currently preparing research paper for conference submission on mechanistic interpretability approaches to fairness evaluation in clinical AI systems

Freelance Consulting   07/2025 – Present
- Developing ML backend for AI-enhanced strategic decision-making platform using historical military case studies
- Built interactive web dashboard (Vercel deployment) for strategic simulation with parameter weighting and visualization
- Implementing hybrid AI architecture combining generative AI preprocessing with custom inference models
- Creating educational interface that processes unstructured historical data to enable users to simulate strategic decisions with contextual analysis and empirically-grounded AI feedback

Technology Optimisation Intern, Windsor Fellowship,   06/2024 – 08/2024
Elsevier - Global Health Markets
- Produced the first Software Development Lifecycle plan for Elsevier's Health Markets, using Confluence and Jira APIs (**React & Node.js**), reducing average time spent on admin-related tasks by 58%.
- Piloted research into quantifying and mitigating racial and gender bias in the Clinical Key AI tool, producing a proof-of-concept **custom retrieval-augmented generation** system for 1,000,000+ peer-reviewed medical papers.

Consultant Computational Neuroscientist, *Neuro-Bio Ltd*   05/2023 – 05/2024
- Conducted independent research on 90+ polyclonal antibodies via **Western Blot** and **ELISA**, producing a database detailing their specificity, saving £500,000 pa on low-quality assays and halving the time taken to reach a lead candidate.
- Co-led an anti-ageing research group with Unilever, using **Python** and **R** for high-throughput compound selection and analysis.
- Ran **cell viability** and **calcium assays** on novel peptides based on gene set enrichment analysis (**GSEA**) and genome-wide association studies (**GWAS**).
- Developed custom workflows, including multivariate cellular response processing with **PyTorch**, combining **DeepCAD** for denoising and **CICADA** for automated activity detection, reducing manual analysis time by 85% and improving data analysis accuracy to 99.98% (from 71%).
- Led in-silico drug development for Neuro-Bio's novel Alzheimer's therapeutic, utilizing both industry-standard (**AutoDock Vina, Schrödinger, GROMACS**) and AI/ML tools (**RoseTTAFold All-Atom, GNNs**) for MD simulations. Resulted in a 40% improvement in the accuracy of binding affinity predictions and 25% better target selectivity using QSAR-optimized structures.
- Produced SOP for in-house ADMET modelling, combining cheminformatics (**RDKit, MOE**), quantum mechanics (**QSite**), and deep learning (**DeepChem**), saving £150k+ on outsourcing and consultancy.

Research Assistant, *Mouthy Ltd* 06/2022 – 12/2023
- Produced an integrated database for clinical market and investor research, resulting in successful early-stage VC funding.
- Identified key barriers to dental insurance access through demographic analysis, revealing healthcare disparities across socioeconomic groups.

🎖 AWARDS

Full Master's Scholarship, *Snowdon Trust*
- Scholarship for exceptional students to enable further study in their field of choice and promote inclusivity and change as leaders in their subject area
- Selected from 572 applicants (top 1.4%)
- Currently developing disability advocacy initiatives through the Snowdon network and Disability Action Research Kollective
- Leading research project investigating the link between hypermobility/EDS and neurodiversity, examining current research landscape and building consensus on diagnostic overlaps
- Snowdon Trust: https://www.snowdontrust.org.uk/ 🔗
- Disability Action Research Kollective: https://libcom.org/tags/disability-action-research-kollective 🔗

AI and Digital Health Hackathon, Overall Winner, *University College London*
- Developed ScleroDx: AI-powered diagnostic tool for systemic sclerosis addressing health equity in rare disease diagnosis
- Implemented GARNN architecture achieving 79% test accuracy on EUSTAR synthetic dataset (13,514 time-series records) with interpretable clinical predictions
- Live deployment: https://scleroderma-api.onrender.com/ 🔗
- See GitHub for code and more info: github.com/Amelia3141/Hackathon 🔗

PLUS Award, *University of Bristol*

Achieved UoB's employability 'PLUS' award for over 70 hours of work experience, volunteering, internships, workshops, and courses

Student Research Festival, First Place

My paper and conference presentation on 'The Neurobiology of Ethical Decision Making' won first place in the Sustainability and Wellbeing Category of Bristol's Student Research Festival.

📁 PROJECTS

3rd Year UG thesis on "Ivabradine and the Risk of Torsades de Pointes: A Case-Based Analysis"
- Achieved a grade of 84 overall and 89 on Viva exam
- Developed BioBERT-based text mining system for automated pharmacovigilance with custom entity recognition
- Created open-access web tool for automatic systematic reviews of torsadogenic medication
- Manuscript in preparation for publication
- GitHub: https://github.com/Amelia3141/IvabradineTdP 🔗

🧠 SKILLS

Programming Languages & Frameworks – Python (Pandas, NumPy, SciPy, PyTorch, Django, scikit-learn) | JavaScript (React, Node.js, Full-stack development) | R (Statistical analysis, ggplot2, tidyverse) | MATLAB | Version Control: Git, GitHub

Data Science & ML – Machine Learning: BERT/BioBERT, RAG models, Knowledge Graphs, GNNs, GARNN (Graph Attention RNN), XGBoost, Random Forest | Mechanistic Interpretability: Sparse Autoencoders (SAEs), Gemma Scope, feature clamping techniques | Statistical Analysis: IBM SPSS, Advanced statistical modeling, SMOTE for imbalanced data | Bioinformatics: GSEA, GWAS analysis | Time-series analysis and temporal modeling for disease progression | High-throughput data processing, web scraping (BeautifulSoup, Selenium), and large-scale data collection

Scientific Computing & Tools – - Molecular Modeling: AutoDock Vina, RoseTTAFold, GNNs | - Image Analysis: ImageJ/Fiji, | - Cheminformatics: RDKit, MOE, QSite

🌐 LANGUAGES

English	● ● ● ● ●	Farsi	● ● ● ● ●
French	● ● ● ○ ○	Italian and Turkish	● ● ○ ○ ○