

Anabel Yong

ucabyon@ucl.ac.uk — 07845636286 — Personal Website: <https://anabelyong.github.io/>

Education

MSc Computational Statistics and Machine Learning, UCL (Distinction, Dean's List 2024) 2024
Modules: Probabilistic and Unsupervised Learning (Gatsby PhD module) (91%), ML Seminar (Marc Deisenroth- DeepMind Chair in AI (88%), Numerical Optimization in Machine Learning (81%)

- **Statistical Natural Language Processing Research @ UCL x META Facebook AI Research**
 - Evaluating performance of DoLA decoding by contrastive layers strategy of Large Language Models GPT, LLAMA7B, FLAN-T5) implemented for instruction following to reduce hallucinations in LLMs

BSc Mathematics & Biology, University of Edinburgh (1st Class Honours) 2019 - 2023
Modules: Bioinformatics (95%), Machine Learning (80%), Numerical Optimisation, Data Structures & Algorithms

- **Thesis (80%): Optimising SOTA data pre-processing pipeline pgFDR for MS-based proteomics data**
 - Analysed false discovery rate (FDR) strategies to detect microproteins in human proteome, thesis converted to PHD proposal for Biomedical AI CDT program and publication to NatureCommunications under review.

Work Experience

- **ML Engineer Intern, Ignotalabs.AI, Austin Tripp, University of Cambridge & Brooks Paige, UCL** June 2024-Sept 2024
 - Developed a novel Gaussian Process Bayesian Optimization algorithm modelling each molecular property independently using Tanimoto Kernel GPs (submission to ICLR Conference 2025).
 - Developed a Kernel-only Gaussian Process Python Package (KERN-GP)
 - Researched Tanimoto Kernel methods to handle molecular fingerprints and multi-objective bayesian optimization methods
- **Data Scientist, Graphcore.AI** June 2022 - Sept 2022
 - Explored research in on-chip and off-chip, GPU optimization techniques (warp-shuffle functions, coalesced access and kernel fusion)
- **Computational Biologist, Kustatscher Lab, University of Edinburgh x Technical University of Munich** Sept 2022 - July 2023
 - Optimised state-of-the-art data pre-processing pipeline, (Link to pgFDR), maximizing human protein detection by 58%, which let to discovering 7000+ novel microproteins
 - Deployed other false discovery rate estimation approaches on human proteome dataset, ProteomicsDB.
 - Analysed performance of pgFDR tool with other SOTA methods in computational proteomics.
- **Statistical Geneticist, UCL Centre of Computational Biology, UCL** March 2022 - May 2022
 - Implemented Bayesian MCMC algorithm to sample probability distributions in p, q, and r frequencies, for ABO Blood Frequency Modelling in C#
- **Statistical Biologist, Department of Systems & Synthetic Biology, University of Edinburgh** June 2020 - Jan 2022
 - Developed stochastic models (stochastic simulation algorithm) & deterministic models to analyse and predict complex data patterns in mRNA decay biochemical kinetics.

Positions Held & Achievements

Vice President, EdIntelligence - AI & Data Science Society Sept 2021 - May 2022
Treasurer, Computer Science Society, University of Edinburgh Sept 2020 - May 2021
Top Scorer in Biological Chemistry 2 May 2021
Top Scorer in Structures and Functions of Proteins 3 Award, University of Edinburgh May 2022

Personal Projects (on Portfolio)

- **Review of Meta-Learning with Shared Amortized Variational Inference**
 - Implemented MAML, Shared AMortized VARIational inference (SAMOVAR), and versatile amortization network (VERSA) meta-learning models to evaluate model's performance on 5-way classification task on miniImageNet dataset.

Skills

Programming: git, linux, R, Python, MATLAB, SQL, Rust, C#, CUDA, Bash (linux shell scripting), LaTeX
Bioinformatics Tools: BLAST, PantherDB, Biopython, Bioconductor, GSEAPy, MaxQuant, Percolator, Fragpipe
Packages: Botorch, Gpytorch, PyTorch, Tensorflow, SentenceTransformers, Numpy, Matplotlib, Pandas