

Anabel Yong

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

Education

MSc Computational Statistics and Machine Learning, UCL, Expected: Distinction 2023 - 2024
Modules: Probabilistic and Unsupervised Learning (Gatsby PHD), Statistical NLP (Facebook AI Research, ML Seminar (Marc Deisenroth- DeepMind Chair))

- **Masters Thesis with IgnotaLabs.AI & Brooks Paige, UCL**
 - Investigating Latent Space Optimization in Deep Generative Models (Grammar Variational Autoencoders) for high-efficacy molecules, and avoiding high toxicity
- **Statistical Natural Language Processing Research @ UCL x META Facebook AI Research**
 - Exploring performance of DoLA decoding strategy on Large Language Models for instruction following

BSc Mathematics & Biochemistry, University of Edinburgh, First Class Honours 2019 - 2023
Modules: Bioinformatics (95%), Introductory Machine Learning (80%), Statistical Biology

- **Research 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 programme at University of Edinburgh

Work Experience

- **Data Scientist, Graphcore** June 2022 - Sept 2022
 - Implemented A/B testing models, (multivariate A/B tests), to compare and evaluate different IPU- powered generative models tailored to dynamics of consumer markets for laboratory automation.
- **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
- **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
 - Organising committee for ClimateHackAI datathon where participants develop machine learning models for data provided by OpenClimateFix to reduce industrial carbon emissions
- **Treasurer, Computer Science Society, University of Edinburgh** Sept 2020 - May 2021
 - Organising committee of HACKTHEBURGH 2021 (largest hackathon in Scotland, sponsored by Google, Optiver and MajorLeagueHacking)
- **Structures and Functions of Proteins 3 Award, University of Edinburgh** May 2022
 - Highest overall mark in my university cohort for SFP3 in biophysical, & experimental methods component (81%)
- **Biological Chemistry 2 Award, University of Edinburgh** May 2021
 - Highest overall mark in my second year cohort for quantitative methods in biochemistry (96%)

Skills

Languages:	R, Python, MATLAB, Haskell, SQL, C#
Development Tools:	Git, VS Code, PyCharm, Google Cloud Platform, Jupyter Notebook
Data Repositories:	NCBI, UniProt, PDB, dbSNP, SFARI, ClinVar, BioBank, PantherDB
Database Management Tools:	MongoDB, MySQL, Neo4j