# Anabel Yong

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


## Summary


MSc in Computational Statistics and Machine Learning from UCL, with a specialization in scalable Bayesian optimization for molecular design. Experienced in developing hardware-optimized ML models using PyTorch and JAX and efficient compute frameworks for molecular property prediction. Strong interest in hardware-aware AI algorithms, distributed ML, and efficient compute for drug discovery applications.


## Publications

- **Yong, A., Paige, B., Tripp, A. (2024). Beyond Scalarization: Extending GP-BO to Pareto Optimization with Independent Tanimoto Kernel Gaussian Processes.** ***(Submitted to ICLR 2025)***

## 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 NLP

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

## Work Experience

- **ML Engineer Intern, IgnotaLabs.AI, Austin Tripp, University of Cambridge & Brooks Paige, UCL** June 2024- Sept 2024
  - Developed a multi-objective Bayesian optimization algorithm (GP-MOBO) using Tanimoto Kernels for small molecule optimization.
  - Deployed a Python package (KERN-GP) for efficient kernel-based Gaussian Processes, enabling scalable, GPU-optimized molecular property modeling.
  - Built scalable ML pipelines with PyTorch, optimizing molecular simulations by 20% through distributed training
- **Research Assistant, Kustatscher Lab, University of Edinburgh x Technical University of Munich** Sept 2022 - July 2023
  - Optimised data pre-processing pipeline (published to PyPI), (Link to pgFDR) to improve protein detection accuracy, identifying over 7000 novel microproteins for molecular analysis.
- **Statistical Genetics Intern, ZiHeng Yang Lab, UCL Centre of Computational Biology, UCL** March 2022 - May 2022
  - - Implemented Bayesian MCMC algorithms to model ABO blood type frequencies, contributing to personalized medicine approaches by estimating key genetic parameters for population studies.
- **Research Assistant, Grima Lab, Department of Systems & Synthetic Biology, University of Edinburgh** June 2020 - Jan 2022
  - Implemented stochastic and deterministic models to predict mRNA decay kinetics, aiding in the understanding of gene regulation mechanisms and supporting drug efficacy studies in preclinical research.

## Personal Projects (on Portfolio)

- **DoLa-based Decoding for Instruction-Following LLMs**
  - Evaluated and adapted DoLa contrastive decoding in T5 and FLAN-T5, enhancing instruction accuracy and reducing hallucinations for large language models (LLMs).

## Skills

**Programming:** Python, C++, R, MATLAB, SQL, git, Linux
**Machine Learning:** PyTorch, JAX, TensorFlow, Botorch, GPytorch
**Bioinformatics/Cheminformatics:** RDKit, PyMOL, UniProt, MaxQuant
**Cloud Computing & HPC:** Google Cloud Platform (GCP), AWServices (AWS)