

# Maciej Wiatrak



macwiatrak@gmail.com •



Cambridge, UK •

<https://macwiatrak.github.io/> •

[github.com/macwiatrak](https://github.com/macwiatrak)

## EDUCATION

### University of Cambridge

*PhD in Machine Learning for Biology, Cambridge Centre for AI in Medicine.*

Cambridge, UK

10/2022 – 10/2026

- Built a large-scale generative model for whole-genome analysis of bacteria. Applied the model to various tasks such as strain clustering, protein function annotation and more showcasing its state-of-art performance and applicability.
- Designed and developed a state-of-the-art ML framework for predicting antibiotic resistance from DNA sequence incorporating gene regulation ([link](#)). Successfully applied it to multiple tasks, including variant scoring and gene expression prediction
- Supervisors: **Prof. Andres Floto**, **Prof Sarah Teichmann**, **Prof. Mihaela van der Schaar**.

### University College London

*Undergraduate, BSc Science & Engineering; Major: Computer Science, Minor: Mathematics.*

London, UK

09/2016 – 05/2019

- **1<sup>st</sup> class honours with distinction** (Dean's List, top 5% of the students).
- Undergraduate topic thesis: "Governing the Commons with Multi-Agent Reinforcement Learning". Advisor: **Prof. Jun Wang**.

## WORK EXPERIENCE

### Genentech

*Machine Learning Intern, Biology Research & AI Department (BRAID)*

San Francisco, CA, US

07/2024 – 12/2024

- Designed and built a model for cell-cell communication analysis.
- Deployed the model to analyse cell-cell interactions dysregulated in the disease.

### BenevolentAI (AI for Drug Discovery)

*Machine Learning Engineer / Team Lead*

London, UK

12/2019 – 10/2022

- Working on target identification and knowledge discovery to predict and validate target hypotheses for drug development.
- Squad lead for applied research team working on data integration using deep learning.
- Developed and deployed a top-performing model for identifying target genes for various diseases and biological mechanisms.
- Proposed and led a project on building a method for providing case-based explainable output from a machine learning model.
- Co-authored 5 papers on graph models and NLP as well as 2 patents.

## SELECTED PUBLICATIONS & PREPRINTS

- **Sequence-based modelling of bacterial genomes enables accurate antibiotic resistance prediction.** *Wiatrak, M., Weimann, A., Dinan, A., Brbić, M., Floto, R. A.* In *bioRxiv*, 2024.
- **Directed Graph Embeddings in Pseudo-Riemannian Manifolds.** *Sim, A., Wiatrak, M., Brayne, A., Paliwal, S., Creed, C.* In *International Conference on Machine Learning (ICML)*, 2021.
- **On Masked Language Models for Contextual Link Prediction.** *Brayne, A., Wiatrak, M., Corneil, D.* In *ACL, DeeLIO*, 2022.
- **Proxy-based Zero-Shot Entity Linking by Effective Candidate Retrieval.** *Wiatrak, M., Arvaniti E., Brayne, A., Vetterle, J., Sim, A.* In *EMNLP, LOUHI (Health text) workshop*, 2022.

## HONORS & AWARDS

- **Best Talk.** "Fighting antibiotic resistance with deep learning". ML in PL conference 2023.
- **4-year GSK & AstraZeneca PhD studentship (2022-2026)**
- **Distinction - Dean's List** (merit-based, awarded to the top 5% of the students); University College London. June 2019.

## METHODS, SKILLS & LANGUAGES

**AI/ML:** Transformers, large-scale training, Graph learning (GNNs & KGE), DNA sequence models, VAEs, pLMs.

**Bioinformatics:** Single-cell genomics (scRNA-seq, scATAC-seq, ChIP-seq, CITE-seq and more), scanpy, sevi-tools.

**Data Engineering:** Docker, Parallel & distributed processing (Spark), ETL pipelines (k8s), Model deployment & serving.

**Frameworks:** PyTorch (advanced), Lightning, HuggingFace, matplotlib, seaborn, Spark (advanced), scikit-learn, AWS, k8s.

**Open source:** Contributor to scverse (tools for single-cell omics), working on DNA-sequence models and data loaders.