Maciej Wiatrak

macwiatrak@gmail.com • Cambridge, UK • https://macwiatrak.github.io/ • github.com/macwiatrak

EDUCATION

University of Cambridge — **Cambridge, UK**
PhD in Machine Learning for Biology, Cambridge Centre for AI in Medicine. 10/2022 – 10/2026

- Building applied ML methods for biology. Focusing on modeling bacterial genomics and protein-protein interactions.
- 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
- Developed a large-scale transformer-based method modelling whole genomes with contextualised protein representations and trained it on 1.3M whole genomes and 3B proteins. Applied it to several tasks, showing SOTA performance.
- Built a protein-protein interaction model for cell–cell communication that integrates sequence, structure and curated prior knowledge.
- Supervisors: **Prof. Andres Floto**, **Prof Sarah Teichmann**, **Prof. Mihaela van der Schaar**.

University College London — **London, UK**
Undergraduate, BASc Science & Engineering; Major: Computer Science, Minor: Mathematics. 09/2016 – 05/2019

- **1st class honours with distinction** (Dean's List, top 5% of the students).
- Undergraduate thesis topic: "Governing the Commons with Multi-Agent Reinforcement Learning". Advisors: **Dr Manuela Dal Borgo** and **Prof. Jun Wang**.

WORK EXPERIENCE

Genentech — **San Francisco, CA, US**
Machine Learning Intern, BRAID Department 07/2024 – 12/2024

- Designed and built a large transformer-based machine learning model for cell-cell communication analysis.
- Deployed the model to analyse cell-cell interactions dysregulated in the disease.

BenevolentAI — **London, UK**
Team Lead / Machine Learning Engineer 04/2022 – 10/2022

- Squad lead for applied research team of 5 people working on target identification.
- Led the development of a framework integrating multiple modalities with deep learning.

Machine Learning Engineer 12/2019 – 03/2022

- Working on target identification and knowledge discovery to predict and validate target hypotheses for drug development.
- 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.

University of Edinburgh — **Edinburgh, UK**
Team Lead / Machine Learning Engineer 09/2019 – 11/2019

- Working on stabilising Generative Adversarial Networks (GANs) using multi-agent reinforcement learning.
- Conducted experiments and wrote a comprehensive survey on the state of Generative Adversarial Network training stabilisation.
- Supervised by Dr Stefano V. Albrecht, Head of the Autonomous Agents Research Group.

Growbots Inc. — **London, UK / Warsaw, Poland**
Account Executive 06/2016 – 11/2018

- Led end-to-end sales cycle for a B2B SaaS platform, including prospecting, product demos, contract negotiation, and onboarding.
- Acquired and closed over 100 new clients across SMB and mid-market segments, consistently exceeding monthly revenue targets.
- Scaled with the company from ~20 to 100+ employees, adapting sales strategies in a fast-growth environment.

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.
- **Simple hierarchical multi-task neural end-to-end entity linking for biomedical text.** *Wiatrak, M.,* Iso-Sipila, J. In EMNLP, LOUHI (Health text) workshop, 2020.
- **Pseudo-Riemannian Embedding Models for Multi-Relational Graph Representations.** Paliwal S., Brayne A., Fabian, B., *Wiatrak, M.*, Sim A.. Accepted to AKBC 2022.

HONORS & AWARDS

- **50 Years VC Fellow.** Part of the 1st 5050 UK cohort supporting deep tech startup founders. 2025.
- **4-year GSK & AstraZeneca PhD studentship (2022-2026)**
- **Dean's List (merit-based, awarded to top 5% of the students); University College London. June 2019.**
- **Best Talk – People's choice award. Science Polish Perspectives (SPP) conference. University of Cambridge. November 2019.**
- **Best Talk Runner-up.** "Fighting antibiotic resistance with deep learning". ML in PL conference 2023.
- **Provost's Scholarship (merit-based, top 5% of the admitted students). Warsaw School of Economics. January 2016.**
- **National Bank of Poland Scholarship for outstanding academic performance (top 1%). September 2015.**
- **Geography Olympian (top 1% of students worldwide), awarded twice. April 2015, April 2014.**

METHODS, SKILLS & LANGUAGES

AI/ML: Large-scale training (DeepSpeed, mixed-precision, multi-node training), Protein & DNA LMs, LLMs, Transformers & LLMs (pretraining, finetuning, prompting), Interpretability (saliency maps, DeepLIFT, GNN Explainer), Graph ML (GNNs, KGEs), Generative modelling (VAEs), contrastive & self-supervised learning
Bioinformatics: Bacterial genomics, Single-cell genomics (scRNA-seq, scATAC-seq, CITE-seq, ChIP-seq, spatial transcriptomics), scanpy, scvi-tools, protein-protein interactions, cell-cell communication.
Data Engineering: Docker, Parallel & distributed processing (Apache spark), ETL pipelines (k8s), Model deployment & serving, FAISS, CI/CD.
Frameworks: PyTorch (advanced), Lightning, HuggingFace, Apache spark (advanced), scikit-learn, AWS, k8s.