

Languages & Skills:

Areas: Full-Stack, Data Engineering, Bioinformatics, Data Analysis, Cloud Compute Infrastructure, Machine Learning

Languages: Javascript (NodeJS, ReactJS, NextJS), Python, C++, SQL, Nextflow (Viash), Git, Cypher, R, Java

Infrastructure: AWS, Azure, Kubernetes, TileDB, Neo4J, SQL, CI/CD

Education

Imperial College London

Medical Bioscience (BSc) 2:1
2018 – 2021

- Relevant modules:
 - Statistics (R)
 - Neuroscience
 - Genetics & Genomics
 - Pharmacology
- Two year-long undergraduate Lab-based research projects on Alzheimer's and Paclitaxel in MCF-7 Cells.
- Undergraduate Machine Learning Researcher with the National Heart and Lung Institute based in Hammersmith Hospital

Charterhouse School

2012 – 2017

- **Higher Level:** Chemistry, History, and Biology
- **Standard Level:** English, Mathematics and German

Online Courses, Hackathons, and Interests

Coursera:

Neural Networks and Deep Learning, Improving Deep Neural Networks: Hyperparameter tuning, Regularisation, and Optimisation, Structuring Machine Learning Projects, Convolutional Neural Networks, Sequence Models, Nand2Tetris, Computational Neuroscience, The Brain and Space, Philosophy and the Sciences: Introduction to the Philosophy of Cognitive Sciences

Hackathons:

IC HealthHack '19: Built a wearable posture monitoring device to dynamically analyse posture and prevent spinal injury.
IC HealthHack'20: Category runner-up. Developed a mental health app which uses ML to calibrate self-affirmation.

Spoken Languages

English (Native)
German (Rudimentary)

Sport

Muay Thai Fighter (CMA; 2022-2024)
Imperial College London: Water Polo BUCS 1st Team (Men's Premier League; 2019-2021).
Imperial College Swimming and Water Polo Club Treasurer (2019-2020)
Berlin Aquaholics (2017-2018)

Work Experience

5/2023-Present **Software Engineer, CompCore / ClinData, insitro, San Francisco** (Remote, FT)
Develop cloud compute infrastructure (K8s, Terraform, AWS) to support software across insitro's scientific and computational domains. Build full-stack omics and embeddings data explorers. Engineer data structures to consolidate data infrastructure and support omics analysis and ML at scale. Produced analyses and reports for target screening with scientific teams. Stood up image viewing infrastructure and tooling for clinical datasets.

12/2021-5/2023 **Lead Data Engineer, Data Science Platform, CZ Biohub, San Francisco** (FT)
Developed a platform (Datahub) to support large-scale research projects by modelling highly heterogeneous lab metadata with graph databases (Neo4J) and setting up a lightweight LIMS system (NextJS/NodeJS). Built pipelines for the mass spectrometry and genomics platforms (OpenPipelines). Communicated with leadership, and users, and coordinated software engineers to integrate related software. Researched graph databases as a means of modelling omics data. Provided ad-hoc analysis for transcriptomics and spatial research.

7/2021 – 10/2021 **Bioinformatics Internship, Illumina, Cambridge** (Remote, FT)
Worked as a bioinformatician for Illumina's EMEA office, within their Population Genomics (PopGen) group. Contributed to methods for merging and saving terabytes of variant called files (VCFs) into a single flat file store (gVCF). Developed test software for triaging bugs and identifying algorithmic slowdowns. Worked on the DRAGEN platform in C++ and Python.

6/2020 – 10/2020 **Bioinformatics Intern, Global Gene Corp, Cambridge** (Remote, FT)
Developed an exome analysis pipeline (nextflow) for variant calling and evaluated called variants using bioinformatics tools and Exomiser in an HPC for the 1 Million Genomes Project. Built a website which enabled geneticists to review and autogenerate reports on likely causative variants for the subject's disease phenotype.

7/2019 – 4/2020 **Software Developer, Automat, Berlin** (Remote, PT/FT)
Automat develops turnkey software solutions. Worked within Automat's 'Workshop Mode' and was responsible for the infrastructure behind the communication software pricing API and a web-scraping tool in AWS. Worked with Docker, SQL, and Lambda.

3/2018- 8/2018 **Gap Year Research and Development Intern, i2x, Berlin** (FT)
i2x builds AI-driven sales generation tools. Responsible for researching and development of a hardware device to support i2x's call assistance service. Fed audio from hardware to i2x's ML algorithms. Designed a website in HTML and Javascript to display results.

9/2017 – 3/2018 **Gap Year Engineering Intern, Leonyte Biosystems, Berlin** (FT)
Leonyte was a small early stage startup aiming to provide real-time testing for pathogens in food. Developed a prototype portable bacterial detection device from research done by UT Austin. Wrote python scripts for visualising biological and system data for presentation and troubleshooting. Smoothed signals and fine-tuned detection algorithms to improve pathogen detection in the embedded C++ software.

Research Experience

Machine Learning Researcher - National Heart and Lung Institute *October 2020 – April 2021*
Developed machine learning models under the guidance of Professor Darrel Francis and Dr. James Howard in the Cardiology Department of Hammersmith Hospital, with the end goal of producing a CMRI (cardiac magnetic resonance imaging) pipeline for automated diagnostic of cardiovascular diseases. Developed a classifier that identified the 'view' of CMRIs, and retrospectively assessed model performance using several metrics. Investigated a potential novel transformer methodology for image data.

Researcher - Imperial College London *October 2018 – March 2019*
This project was an investigation into the effect of adiponectin on β -Amyloid secretion in PDK1-Knockdown HEK293 cells, as a potential molecular target for treating Alzheimers. I worked with a group of four other researchers to develop knockdown cells with CRISPR, and evaluate the impact of adiponectin on the β -Amyloid secretion pathway through protein and genomic expression.