LINFENG WANG

wanglinfeng1115@gmail.com ◇ +44 7599610897 ◇ linkedin.com/in/w15 ◇ linfeng-wang.github.io/

TECHNICAL STRENGTH

Language & Frameworks:	**Python, PyTorch**, PyG, R, Bash, MySQL, HTML/CSS, C++
Machine learning:	Scikit-learn Transformers, LLMs, CNNs, RNNs, GNNs, VAEs, XGBoost
Development Tool:	Pandas, NumPy, Docker, Git, Nextflow, Google Cloud, Jupyter
Visualization:	OpenCV, Unity , Matplotlib, Plotly, ggplot2

EDUCATION

London School of Hygiene and Tropical Medicine , DPhil Computational Genomics   **Oct 2021 – Oct 2025 (expected)**
Dissertation: *Machine Learning-Enhanced Drug Resistance and Bioinformatics Transmission Profiling of Tuberculosis Using Genome Sequencing*

Imperial College London, MRes Bioengineering (Hons), Merit   **Oct 2019 – Oct 2020**
Modules: Computational & Statistical Methods for Research, Frontiers in Bioengineering, Biomaterials
Dissertation: *Design of an Artificial Bruch's Membrane from Synthetic Polyesters.*

King's College London, BSc Biochemistry (Hons), 1st   **Sep 2016 – Jul 2019**
Module selected: Bioinformatics, Protein structure and design, Human genomics.
Dissertation: *Investigation of Concordance Between Molecular Dynamics Simulation and FRET Biosensor using Designed Protein Linker System.*

EXPERIENCE

PhD researcher - LSHTM (Clark Campino Phelan lab), London UK   **Jul 2022 – Sept 2025**
Project: Deep Learning and Statistical Modelling for Tuberculosis Drug Resistance and Transmission
- Developed interpretable deep learning models (**CNN, RNN, GNN, Transforme**r) in **PyTorch and JAX** to predict TB drug resistance from genomic data
- Built **statistical pipelines** in Python for GWAS, PCA, logistic regression, and odds ratio analysis on Philippine TB datasets
- Created a genetic algorithm-based tool for gene amplicon design using **Python and Nextflow**
- Reduced model complexity and compute time by transitioning from raw sequence to SNP frequency input

Machine learning consultant - Deep Science Venture, London UK   **Mar 2025 -**
Project: Sequence-Based Drug Discovery using Deep Generative Models
- Trained **CNN, RNN, and VAE** models on biological sequences with data augmentation and hyperparameter optimization
- Applied **SHAP, LIME, and DeepLIFT** to extract interpretable biological insights from deep models
- Supported strategic model design for generative sequence discovery tasks

Data study group hackathon participant – The Alan Turing Institute, London UK.   **Jan 2024**
Project: Shallow Gas Hazard Detection in Offshore Seismic Data
- Led deep learning model development for detection and segmentation on geophysical imagery using **CNNs with contrastive learning**
- Delivered 90% classification accuracy on legacy seismic datasets under 2-week sprint
- Collaborated across domain boundaries to translate geoscience challenges **into ML solutions**

Machine learning Intern - Linkgevity, London UK   **Aug 2024 – Oct 2024**
Project: Graph-Based Drug-Drug Interaction Prediction
- Built and optimized **MLP and GNN models** in Python for predicting compound interactions
- Conducted benchmarking studies across model architectures to improve predictive performance
- Scaled ML training workflows using Google Cloud Platform for **high-throughput experimentation**

Data Science intern - ByteDance, London UK   **Aug 2022 – Feb 2023**
Project: Neoantigen Insight Platform and Market Landscape Scouting
- Built **SQL-integrated chemical database** and applied autoencoders to reveal neoantigen patterns across biological datasets
- Coordinated **interdisciplinary market research** on biologics and small molecules
- Authored bi-weekly **research briefings** and developed end-to-end ML pipelines for **discovery support**

PUBLICATIONS (6 First authored, 9 in total)

- *A novel tool for designing targeted gene amplicons and an optimised set of primers for high- throughput sequencing in tuberculosis genomic studies, BioRxiv, (submitted for publication).*
- **Website:** https://genomics.lshtm.ac.uk/webtoast/#/ **software package**: https://pypi.org/project/toast-amplicon/
- *LSTM-Based Transfer Learning Models for Tuberculosis-Targeted Antimicrobial Peptide Classification and Generation*. (*submitted for publication)*
- *Decoding Positive Selection in Mycobacterium tuberculosis with Phylogeny-Guided Graph Attention Models*. (forth coming).
- Data Study Group Final Report: British Geological Survey - Detecting Shallow Gas from Marine Seismic Images". The Alan Turing Institute (2025)