

王临风 Linfeng Wang, PhD

wanglinfeng1115@gmail.com · +44 7599 610897 · linkedin.com/in/w15 · linfeng-wang.github.io

专业技能

编程语言与框架	Python, PyTorch, PyG, R, Bash, MySQL, HTML/CSS, C++
机器学习	Scikit-learn, Transformers, LLMs, CNNs, RNNs, GNNs, VAEs, XGBoost
开发工具	AWS, GCP, Azure, Docker, Git, Nextflow, Pandas, NumPy, Jupyter, Hugging Face, LangChain, Numba
可视化	Matplotlib, Plotly, ggplot2, Streamlit, Flask

工作与科研经历

机器学习工程师 — Biographica

2026年4月 — 至今

英国伦敦

- 开发、测试并评估基于 DNA 序列与生物学背景预测基因表达的深度学习模型。
- 在 Azure 与 AWS 平台上搭建可扩展的 PyTorch 训练与微调流程，涵盖数据集构建、模型评估、超参数优化与可复现实验。
- 将机器学习应用于基因编辑与作物性状预测，整合多组学数据。

博士研究生 — 伦敦卫生与热带医学院

2021年10月 — 2026年3月

Clark-Campino-Phelan 实验室，英国伦敦

- 使用 PyTorch 开发容器化 (Docker)、可复现的深度学习模型 (CNN、RNN、GNN、Transformer)。
- 在高性能计算集群上用 Python 构建统计分析流程，对高维数据进行 GWAS、PCA、逻辑回归与优势比分析。
- 用 Python 开发网页应用与后端软件 (PyPI 工具包)，实现扩增子引物设计的自动化。
- 开发基于 RNN 的生成式模型，用于新型抗菌肽序列的设计与分类，准确率达 91%。
- 开发基于 GNN 的分类模型，依据系统发育树预测自然选择，准确率达 88%。
- 构建并部署整合 ChromaDB、LangChain 与 OpenAI API 的 RAG 大模型问答机器人，用于课程内容检索与反馈收集。

机器学习顾问 — Deep Science Venture

2025年3月 — 2025年10月

英国伦敦

项目：基于深度生成模型的序列药物发现

- 主导 CNN、RNN 与 VAE 模型在生物序列上的开发，结合数据增强与超参数优化。
- 构建基因表达预测模型，准确率达 86%。
- 应用 SHAP、LIME 与 DeepLIFT，从深度模型中提取可解释的生物学信息。
- 主导生成式序列发现任务的模型设计方向。

机器学习实习生 — LinkGevity

2024年8月 — 2024年10月

英国伦敦

项目：基于图神经网络的药物相互作用预测

- 复现并改进 GNN 模型，整合 BERT 预测药物化合物间的相互作用。
- 训练药物相互作用模型用于药物设计，预测性能提升 15%。
- 利用 Google Cloud Platform 与 AWS 扩展机器学习训练流程，支持高通量实验。

Data Study Group 参与者 — 艾伦·图灵研究所

2024年1月

英国伦敦

项目：海上地震勘探数据中的浅层气体隐患检测

- 主导深度学习模型开发，使用结合对比学习的 CNN 完成地球物理图像的检测与分割。

- 在两周内于历史地震数据集上实现 93% 的分类准确率。
- 跨学科协作，将地球科学问题转化为机器学习解决方案。

数据科学实习生 — 字节跳动

2022年8月 — 2023年2月

英国伦敦

项目：新抗原洞察平台与市场调研

- 构建集成 SQL 的候选药物化学数据库。
- 使用自编码器挖掘生物数据集中免疫相关分子的潜在规律。
- 协调生物制剂与小分子方向的跨学科市场调研，并撰写双周研究简报。
- 开发面向药物发现的端到端机器学习流程。

代表性论文

- Wang et al. **TOAST: A novel tool for designing targeted gene amplicons.** *BMC Genomics* (2025). doi:10.1186/s12864-025-12247-9
- Wang et al. **LSTM-Based Models for Antimicrobial Peptide Classification and Generation.** *Bioinformatics Advances* (2025). doi:10.1093/bioadv/vbaf274
- Wang et al. **Decoding Positive Selection with Phylogeny-Guided Graph Attention Models.** *BMC Bioinformatics* (2026). doi:10.1186/s12859-026-06583-0
- Wang et al. **ML for stepwise prediction of TB treatment outcomes.** *Research Square* (预印本, 2025) . doi:10.21203/rs.3.rs-7558046/v1
- Data Study Group Report: British Geological Survey — Detecting Shallow Gas from Marine Seismic Images. 艾伦·图灵研究所 (2025).

教育经历

伦敦卫生与热带医学院（伦敦大学）· 计算基因组学 博士

2021年10月 — 2026年3月

论文：结核病耐药性的整合基因组测序与机器学习研究及诊断工具开发

奖学金：UKRI — BBSRC — LiDo 博士奖学金

伦敦帝国理工学院 · 生物工程 研究型硕士（荣誉），优异成绩 Merit

2019年10月 — 2020年10月

相关课程：科研数据分析

伦敦国王学院 · 生物化学 理学学士（荣誉），一等成绩

2016年9月 — 2019年7月

相关课程：生物信息学、蛋白质设计与功能、分子动力学模拟

奖项：Wellcome Trust 生物医学暑期研究奖学金（2018）