莫纳什大学人工智能硕士在读 · 2026 年 12 月毕业

李北辰 Beichen Li

个人简历 · 2026 年 11 月

澳大利亚 墨尔本 me@beichen.li beichen.li github.com/Beicxxxx

正在寻找 2027 年入学的博士职位，以及机器学习研究或工程岗位。

个人概述

我关注的问题是：AI 系统什么时候足够可靠，以及如何诚实地测量这一点。一项研究检验字形生成模型能否以字体自身的风格，画出这款字体中从未出现过的部件（未见部件）——这是一个分布外泛化问题，在分子设计中有直接的对应；另一项研究考察通用机器学习原子间势（基础模型）是否继承训练标签的赝势约定。我本科学习应用化学，也能端到端搭建机器学习系统：从已发布安卓应用中的蒸馏大模型，到横跨多家云平台的 GPU 实验管线。

教育背景

人工智能硕士 — 莫纳什大学（Monash University）

2025.03 – 2026.12（预计）

澳大利亚 克莱顿

主要课程：深度学习；人工智能基础（搜索、贝叶斯网络、强化学习）；IT 研究与创新方法；数据库系统；项目管理；行业实践项目

应用化学 理学学士 — 兰州大学

2019.09 – 2023.06

中国 兰州

毕业论文：《2-叠氮基-4,4′-联苯二羧酸钕配合物的合成及荧光性质研究》（导师：郭礼荣），研究对象是一种钕配合物（镧系金属有机框架）。全部实验由本人独立完成：四步合成配体，培育单晶并解析其 X 射线结构，以 ^{1}H NMR、PXRD、TGA、IR 表征，并测定近红外荧光

研究经历

中日韩字体生成中的未见部件问题

2026.08 – 至今

独立研究 · 论文撰写中，复现进行中

字体生成模型能否用一款字体自身的风格，画出这款字体中从未出现过的部件？这是把一款字体扩展到数千个缺失字符时的核心难题。

- 设计：6 款目标字体 × 4 类部件（小、日、竹、米），每格 3 个随机种子。每款字体微调两次：一次删去训练字形中含该类部件的全部字，一次保留，然后在同一批留出字上比较生成误差
- 与预训练条件交叉：生成器先在 330 款其他字体上预训练，预训练池含或不含该类部件；随机删除对照则从预训练池里随机删去与该部件无关的字
- 在另一种架构的生成器上做复现，正在进行；分析方案在看到结果前确定

通用机器学习势如何继承赝势约定

2026.11 – 至今

分析计划已于 2026 年 11 月冻结 · 计划 2026 年 12 月发布 arXiv 预印本

- 研究通用势是否继承训练标签的赝势约定，以镱为例：2023 年 5 月，Materials Project 把镱的赝势从 Yb_2 换成了 Yb_3；两种约定分别经由两个数据集进入模型——MPtrj（更换前，Yb_2）与 OMat24（Yb_3）
- 研究对象是公开发布的模型权重：MACE、eqV2、Nequix 同一架构各有 MPtrj 版和 OMat24 版
- 考察约定由网络的哪一部分承载，以及只重训元素嵌入层里镱的那一行，能否把通用势从一种约定移到另一种、同时不改变其他元素的结果，并与重训全部镱相关参数、常规微调作比较
- 代码与数据随投稿公开（MIT / CC BY 4.0）

李北辰

me@beichen.li · beichen.li

应用与工程

Edibli — 过敏原感知食品扫描应用（安卓）

2026.08 – 2026.10

莫纳什 FIT5120 行业实践项目 · AI/大模型负责人、主要贡献者，5 人团队 · 全仓库 622 次提交中的 583 次

- “规则判定、AI 解释”：确定性引擎决定所有涉及安全的过敏原判定，大模型只负责表述，并经过 14 项护栏检查与评测门槛（零矛盾判定、零虚构过敏原）
- 将 Gemma 4 26B 教师模型蒸馏到 Qwen3.5-2B：LoRA 微调（PEFT，r=32，1.4 万条经护栏筛选的样本），量化为 1.3 GB GGUF；留出集护栏通过率 95.5%，配料规范化召回率 97.5%（人工审核）
- 发布 v3.1.0，3,559 项自动化测试通过并经真机验证；Node.js、Oracle 自治数据库与约 480 万条商品的 ETL、Docker、GitHub Actions 发布；开放数据层开源为 openfood-au

IT 与 AI 解决方案助理 — New Asian Supermarket

2025.04 – 2025.11

墨尔本

- 部署商品图像识别模型实现自动商品识别，构建供应商匹配系统和基于销售、利润与库存数据的动态定价算法；维护数据库、内部工具与网络设施

开源项目

霞鹜文楷 — AI 生成字形

2023 – 2024

- 为霞鹜文楷整个字体系列补全 500 多个 AI 生成的字形，逐一对照原字体风格审核；其中繁体版（霞鹜文楷 TC）收录于 Google Fonts

ColorOS-Font-Injector

2026

- 安卓应用（Java、Shizuku）与智能体技能：在主题商店完整性校验与系统拷贝之间换入字体文件，无需 Root 即可把任意 TrueType 字体设为 ColorOS 15/16 系统界面字体；fontTools 管线缩放到 256 UPEM 并保留字距

cross-harness-sync 与 RepoLedger

2026

- 在多台机器间协调大模型智能体长期协作的工具：以 Git 中的 Markdown 保存状态，分层读取上下文、单写者锁、固定交接与分级审查；RepoLedger 提供稳定编号、证据锚点与静态引用检查（256 项测试）

技能

研究方法	分析方案在看到结果前确定、统计功效分析、等价性检验、方差分量、植入阴性盲测、评分者间信度
机器学习	PyTorch、diffusers；从零实现的 GAN（pix2pix）与流匹配生成器；CNN、迁移学习、交叉验证、贝叶斯网络、强化学习（价值迭代、Q-learning）
大模型	LoRA / PEFT 微调、知识蒸馏、vLLM、llama.cpp（GGUF）、护栏与评测门槛、多智能体编排
AI for Science	PyTorch、mace-torch、fairchem、Nequix、ASE（机器学习原子间势）；合成实验、单晶衍射、NMR、PXRD、TGA、IR
工程	Python、SQL、Java、JavaScript（Node.js）、Dart；Android、Flutter；Git、pytest、GitHub Actions、Docker
算力	字体生成器在 Colab、AWS 与 Nectar Research Cloud 的 NVIDIA L4、L40S、A100 上训练；自建容错作业队列，在云 GPU（A10、A16、T4）与笔记本 RTX 4060 上运行训练与结构弛豫，并做跨 GPU 可复现性核对

其他

语言	毕业时间	推荐人
英语、中文、日语	硕士 2026 年 12 月毕业	可应要求提供