MASTER OF AI, MONASH UNIVERSITY · GRADUATING DEC 2026

Beichen Li 李北辰

Curriculum Vitae · November 2026

Melbourne, Australia me@beichen.li beichen.li github.com/Beicxxxx

Seeking a PhD position starting in 2027, and ML research or engineering roles.

PROFILE

I study when an AI system is reliable enough to use, and how to measure that honestly. One study asks whether glyph-generation models can draw, in a font's own style, components that font has never shown — an out-of-distribution question with direct analogues in molecular design. Another asks whether foundation interatomic potentials inherit the pseudopotential conventions of their training labels. Trained first as a chemist, I also build ML systems end to end, from a distilled LLM in a shipped Android app to GPU pipelines across several clouds.

EDUCATION

Master of Artificial Intelligence — Monash University

Mar 2025 – Dec 2026 (expected)

Clayton, Australia

Coursework: Deep Learning; Fundamentals of AI (search, Bayesian networks, reinforcement learning); IT Research & Innovation Methods; Database Systems; Project Management; Industry Experience Studio

BSc, Applied Chemistry — Lanzhou University

Sep 2019 – Jun 2023

Lanzhou, China

Thesis: *Synthesis and Fluorescence Properties of Neodymium 2-Azido-4,4′-biphenyldicarboxylate Complex* (supervisor: Lirong Guo) — a neodymium complex, a lanthanide metal–organic framework (Ln-MOF). All experimental work performed independently: four-step linker synthesis, single-crystal growth and X-ray structure, ^{1}H NMR, PXRD, TGA and IR characterisation, and near-infrared luminescence

RESEARCH

Component support gaps in CJK font generation

Aug 2026 – present

Independent study · paper in preparation, replication under way

Can a font generator draw, in a font's own style, a component that font has never shown? This is the core obstacle to extending a typeface to thousands of missing characters.

- Design: 6 target fonts × 4 component classes (小, 日, 竹, 米), three random seeds per cell. Each font is fine-tuned twice — once with every training glyph containing the component removed, once with them kept — and generation error is compared on the same held-out characters
- Crossed with pretraining: the generator is first pretrained on 330 other fonts, with or without the component class in the pretraining pool; a random-deletion (placebo) control instead removes unrelated characters from the pool
- Replication on a generator with a different architecture is under way; analysis plan fixed before results

Pseudopotential conventions in foundation interatomic potentials

Nov 2026 – present

Analysis plan frozen Nov 2026 · arXiv preprint planned Dec 2026

- Asks whether foundation interatomic potentials inherit the pseudopotential convention of their training labels. Test case: in May 2023 the Materials Project replaced its ytterbium pseudopotential Yb_2 with Yb_3; the two conventions reach models through two datasets, MPtrj (before the change, Yb_2) and OMat24 (Yb_3)
- Studies publicly released weights: MACE, eqV2 and Nequix each have an MPtrj version and an OMat24 version of the same architecture
- Asks which part of the network carries the convention, and whether retraining only ytterbium's row in the element-embedding layer moves a model from one convention to the other while leaving other elements' results unchanged, compared with retraining all ytterbium-related parameters and with standard fine-tuning
- Code and data to be released with the submission (MIT / CC BY 4.0)

Beichen Li

me@beichen.li · beichen.li

ENGINEERING

Edibli — allergen-aware food-label scanner (Android)

Aug – Oct 2026

Monash FIT5120 Industry Experience · AI/LLM lead and primary contributor, team of 5 · 583 of 622 commits

- "Rules decide, AI explains": deterministic engines settle every safety-critical allergen verdict; the LLM only rewords it, behind 14 guardrail checks and evaluation gates (zero contradicted verdicts, zero invented allergens)
- Distilled a Gemma 4 26B teacher into Qwen3.5-2B with LoRA (PEFT, r=32, 14k guardrail-filtered examples), quantised to a 1.3 GB GGUF for llama.cpp: 95.5% guardrail pass on held-out summaries, 97.5% audited ingredient-normalisation recall
- Shipped v3.1.0 with 3,559 passing automated tests and real-device checks; Node.js, Oracle Autonomous Database with an ETL over ~4.8M Open Food Facts products, Docker, GitHub Actions releases; open-data layer released as openfood-au

IT & AI Solutions Assistant — New Asian Supermarket

Apr – Nov 2025

Melbourne

- Deployed an image-recognition model for automatic product identification, a supplier-matching system and a dynamic pricing algorithm on sales, profit and inventory data; maintained databases, internal tools and network infrastructure

OPEN SOURCE

LXGW WenKai (霞鹜文楷) — AI-generated glyphs

2023 – 2024

- Completed 500+ AI-generated glyphs for the whole LXGW WenKai font family, each reviewed against the source style; its Traditional Chinese edition (LXGW WenKai TC) is published on Google Fonts

ColorOS-Font-Injector

2026

- Android app (Java, Shizuku) and agent skill that install any TrueType font as the system UI font on ColorOS 15/16 without root, by swapping the file between the Theme Store's integrity check and its system copy; fontTools pipeline rescales to 256 units per em while keeping kerning

cross-harness-sync & RepoLedger

2026

- Tools for running long multi-agent LLM projects across machines with Markdown state in Git: layered context, a single-writer lock, fixed handoffs and tiered review; RepoLedger adds stable IDs, evidence anchors and static reference checks (256 tests)

SKILLS

Research methods	Analysis plans fixed before results, power analysis, equivalence testing, variance components, blind planted-negative tests, inter-rater reliability
Machine learning	PyTorch, diffusers; from-scratch GAN (pix2pix) and flow-matching generators; CNNs, transfer learning, cross-validation, Bayesian networks, reinforcement learning (value iteration, Q-learning)
LLMs	LoRA / PEFT fine-tuning, knowledge distillation, vLLM, llama.cpp (GGUF), guardrails and evaluation gates, multi-agent orchestration
AI for Science	PyTorch, mace-torch, fairchem, Nequix and ASE for machine-learned interatomic potentials; lab synthesis, single-crystal XRD, NMR, PXRD, TGA, IR
Engineering	Python, SQL, Java, JavaScript (Node.js), Dart; Android, Flutter; Git, pytest, GitHub Actions, Docker
Compute	Font generators trained on NVIDIA L4, L40S and A100 across Colab, AWS and Nectar Research Cloud; a fault-tolerant job queue running training and relaxation jobs across cloud GPUs (A10, A16, T4) and a laptop RTX 4060, with cross-GPU reproducibility checks

OTHER

LANGUAGES

English, Mandarin, Japanese

AVAILABILITY

Master's completes December 2026

REFERENCES

Available on request