

包启明 (Qiming Bao)

LLM Post-Training · Alignment · Logical Reasoning · Agentic RL

bqmbill714@gmail.com | 17826833549 | [个人主页](#) | [GitHub](#) | Google Scholar H-index: 9, 引用 400+

个人简介

博士毕业于 QS 前 100 新西兰奥克兰大学，专注 LLM 后训练 (SFT/DPO/RLHF)、Reward Model 设计与逻辑推理对齐。近期工作 RLearner-LLM 提出 Hybrid-DPO (NLI 信号+Verifier 混合奖励)，消除 verbosity bias，NLI 对齐效果最高提升 6x；Conflict-Aware Fusion 定义并解决 LLM 逻辑惰性 (Logic Inertia)，矛盾场景下推理精度达 100%。AMR-LDA 方法在 ReClor 排行榜第一 (首个隐藏测试集 >90% 团队)。具备 PyTorch/PEFT/Flash-Attention 工程实战能力，有 Agentic workflow、RL 仿真环境与金融 AI Agent 完整项目经验。

教育背景

博士 · 计算机科学 (全额奖学金)	新西兰奥克兰大学 (QS 前 100) 导师: Michael Witbrock 教授	2020.02 – 2025.09
一等荣誉学士 · 计算机科学	新西兰奥克兰大学	2018.07 – 2019.09
学士 · 计算机科学	中国计量大学	2014.09 – 2018.06

核心技能

后训练/对齐	SFT · DPO · Hybrid-DPO · RLHF/PPO · GRPO · Reward Model · NLI-as-reward · 偏好数据合成
推理/数据	逻辑推理 · AMR · 数据增强 · OOD 泛化 · Chain-of-Thought · 多轮决策轨迹合成
训练工程	PyTorch · PEFT/LoRA · Flash-Attention 2 · vLLM · Megatron · int4 量化 · 单卡 7B 训练
Agentic/仿真	LLM Agent · Tool Use · RL 仿真环境 · PPO 奖励对齐 · 金融 AI Agent · 多轮执行轨迹
多模态/文档	Qwen3.5-9B · Gemma 4 · InternVL2 · LayoutLMv3 · ERNIE-LayoutX · OCR · 长序列 4096 扩展

研究与工作经历

大语言模型后训练与逻辑推理研究 (博士) 2020.02 – 2025.09

Strong AI Lab, NAO Institute, 新西兰奥克兰大学

- RLearner-LLM / Hybrid-DPO**: 设计融合 DeBERTa-v3 NLI 信号与 Verifier LLM 的自动化偏好流水线，系统性克服标准偏好信号的 verbosity bias 问题；跨 LLaMA-2-13B / Qwen3-8B / Gemma 4 三架构五领域验证，NLI 对齐最高提升 6x，Qwen3-8B 版本 pairwise 胜率 95%。 ([arXiv 2605.04539](#))
- Conflict-Aware Fusion / Logic Inertia**: 首次定义并量化 LLM 在规则冲突场景下推理归零的失效模式 (Logic Inertia)，提出双过程认知架构将前提验证与逻辑推导解耦；矛盾注入/基础任务双压力测试均达 100% 精度，显著提升缺失证据鲁棒性。 ([arXiv 2512.06393](#))
- AMR-LDA (ReClor 第一)**: 提出 AMR 逻辑等价数据增强框架，结合 GPT-4 Prompt Augmentation，在 NUS 主办的 ReClor (LSAT 题目) 竞赛中排名第一，为首个隐藏测试集超 90% 精度的团队。 ([ACL Findings 2024](#) | [演示](#))
- OOD 逻辑推理系统评估**: 设计四类结构扰动压力测试 (规则删除/矛盾注入/逻辑保持改写/多律等价叠加)，系统评估生成式与判别式 LLM 在分布外逻辑推理的鲁棒性，揭示微小扰动致性能大幅下降；IJCAI 2024 被引超 100 次。 ([IJCAI 2024](#))
- Iterative Enhancement / 教育 LLM**: 开发解释生成与评估模块闭环迭代框架，提升 Learnersourced 平台多项选择题解释质量，AAAI 2025 接受。 ([论文](#))

AI 研究员 / 工程师

2022.07 – 至今

Xtracta (新西兰白名单 AI 软件公司)

- 多模态大模型训练与评测**: 使用 PEFT/LoRA 对 Qwen3.5-9B、Gemma 4 E4B / 31B、InternVL2 进行持续训练与评测；在医疗 PII 脱敏 benchmark (CustodianAI) 上 Qwen3.5-9B 以 65.5 综合分领先 GPT-5 (70.5) 外同尺寸全部开源模型，Gemma 4 E4B 达 73.7；集成 Flash-Attention 2 使 FP16 峰值显存降低 50%，int4 量化实现单张 A4090 (24GB) 完整训练流程。 ([评测平台](#))
- 长序列扩展 (512 – 4096)**: 通过滑动窗口 + Longformer 全局注意力掩码将 LayoutLMv3 / ERNIE-LayoutX 序列长度扩展 8 倍，在 XFUND、FUNSD 及内部数据集的 Token Classification / Relation Extraction 任务上 F1 显著提升，GPU 显存无明显增加。
- LayoutLMv3 多模态预训练复现**: 完整复现微软未开源的 MLM + MIM + WPA 三目标预训练流程，整合 Global Attention Mask 强化文本与视觉向量对齐，具备从零训练能力。
- 仿射变换数据增强 & RDTI 申报**: 应用仿射变换提升文档对齐鲁棒性；成功申请新西兰 Callaghan Innovation RDTI 资助 (2022/2023 年各 15% 研发税收抵免)。

客座副教授 / 合作导师 (AI/NLP/多模态)

2025 – 至今

北京第二外国语学院 (BISU)

- **地理空间救灾 AI 智能体 (GeoDisaster-FM)** : 构建"遥感感知 → 神经符号推理 → RL 决策"完整闭环; 对比 U-Net / DeepLabV3+ / Google AlphaEarth, Sen1Floods11 留一区域 × 多 seed + xBD 留一灾种矩阵消融, 变化检测 F1 +0.087; PPO 学习标注策略, 决策答案与真值相关性 $r=0.971$ (10 个真实灾害事件), 处理 0.03s/chip vs 专家 1–3 天, 目标投稿 Nature Communications。 ([项目演示](#))
- **洪涝灾害预测 (Foundation Model)** : 融合 Sentinel-1/2 多模态遥感数据, 构建可复现实验流水线与实时评测 dashboard (GitHub Pages + Cloudflare Pages 双部署), 覆盖跨区域/跨灾种 10+ 真实事件。
- **短剧多模态翻译系统**: 端到端 pipeline: Qwen3.5-9B / Qwen3-VL / InternVL2 / Gemma 4 字幕识别 (帧率消融 1/2/5fps) → LoRA 微调翻译 → GPT-SoVITS v3/F5-TTS 语音合成; ASR+OCR 自适应融合纠正幻觉, Whisper WER/CER 评估, 论文投稿中。

AI 技术负责人 (顾问)

2025.11 – 至今

Kerrio.ai (新西兰 AI 软件公司)

- **MI 对话训练系统 MVP**: 搭建面向动机式访谈 (Motivational Interviewing) 的大模型对话训练与评估闭环, 涵盖数据准备、提示词模板设计、多轮对话行为约束与一致性优化; 建立"训练 → 评估 → 回归检查"迭代机制, 架构支持扩展为 Agentic 工作流 (工具调用、安全策略、合规模块)。 ([演示](#))

量化交易与 AI Agent 项目 (个人项目)

2025 – 至今

- **AlphaTrader Pro (AI Trading Agent)** : 构建"多源市场感知 → LLM 信号生成 → 风控校验 → 自动执行 → 结果反馈"完整链路; 封装行情查询/指标计算/事件扫描/风险评估/订单执行 Agent 工具链, 对接 Alpaca/Moomoo, 设计 Kelly sizing、VIX 风险缩放、冷却期、止损监控等风控模块。 ([GitHub](#))
- **AuroraBid (RL 广告竞价仿真)** : 基于 RL/Bandit 方法构建广告竞价仿真环境, 支持多策略对比与离线评估, 采用 CI/CD + 自动化测试可复现实验流水线。 ([演示](#))
- **LLM-Accounting (AI 财务自动化)** : 视觉语言模型发票 OCR + 大模型账务 pipeline, 自动生成会计分录与财务报表; 本地 Ollama + Qwen 推理保障数据隐私, Next.js + React 前端。 ([GitHub](#))

NLP 研发实习 · 北大高研院 / 医疗 NLP · Precision Driven Health

2018 – 2020

- 会议机器人 NLP 系统 (摘要提取/文本分割/主题预测); 提出 HBAM 医疗文本相似度模型 (ACSW 2020, 引用 90+), 构建医疗 PII 脱敏 pipeline, 获 AUT Venture 6 万 NZD 投资推动 CustodianAI 成立; CustodianAI 实时评测平台对比 Gemma 4、Qwen3.5-9B、GPT-5 等 10+ 模型在患者隐私保护任务上的表现。 ([论文](#) | [评测平台](#))

代表性论文 (一作 / 共同一作)

[1] Qiming Bao et al.. [RLearner-LLM: Balancing Logical Grounding and Fluency in Large Language Models via Hybrid Direct Preference Optimization](#). arXiv 2605.04539 (2026)

[2] Qiming Bao, Xiaoxuan Fu, Michael Witbrock.. [Conflict-Aware Fusion: Mitigating Logic Inertia in LLMs via Structured Cognitive Priors](#). arXiv 2512.06393 (2025/2026)

[3] Qiming Bao et al.. [Abstract Meaning Representation-Based Logic-Driven Data Augmentation for Logical Reasoning](#). ACL Findings 2024 (CCF-A, Core A*)

[4] Qiming Bao et al.. [Exploring Iterative Enhancement for Improving Learnersourced MCQ Explanations with LLMs](#). AACL 2025 (CCF-A, Core A*)

[5] Gaël Gendron, Qiming Bao et al.. [Large Language Models Are Not Strong Abstract Reasoners](#). IJCAI 2024 (CCF-A, Core A*, 引用 100+)

[6] Nathan Young, Qiming Bao et al.. [AbductionRules: Training Transformers to Explain Unexpected Inputs](#). ACL Findings 2022 (CCF-A, Core A*)

[7] Qiming Bao et al.. [Multi-Step Deductive Reasoning Over Natural Language: An Empirical Study on OOD Generalisation \(Gated Attention cited by Qwen3\)](#). IJCLR-NeSy 2022

[8] Lin Ni, Qiming Bao et al.. [DeepQR: Neural-based Quality Ratings for Learnersourced Multiple-Choice Questions](#). AACL 2022 (CCF-A, Core A*)