吴明昊

minghaowu.self@qq.com
+86 17600442356
https://minghao-wu.github.io/

个人亮点

- 8年以上机器学习、自然语言处理，和人工智能的研究和开发经验。
- 在顶级会议或期刊（如ICML、ACL、EMNLP等）上有优秀的发表记录。
- Google Scholar引用量900+，H指数为13。
- Github Star数累计3.6K+，贡献了多个开源项目。
- 多段顶级机构实习或访问经历。

教育背景

莫纳什大学 (QS 2025 Ranking: 37) *2021年12月 - 2025年12月（预计）*
计算机科学博士
导师：Gholamreza Haffari、George Foster、Lizhen Qu和Trang Vu
研究兴趣：深度学习、自然语言处理、大语言模型、多智能体、多语言、机器翻译

墨尔本大学 (QS 2025 Ranking: 13) *2016年3月 - 2018年7月*
信息技术硕士
导师：Trevor Cohn

悉尼大学 (QS 2025 Ranking: 18) *2013年3月 - 2016年3月*
信息系统理学学士

工作经历

研究实习生 *2024年11月 - 2025年5月*
阿里巴巴集团，中国

- 参与构建覆盖超过200种语言的综合性多语言基准测试。该项目包括收集现有的多语言基准测试，将英语基准测试翻译成其他语言，并在新基准上评估最新的大语言模型。

研究实习生 *2023年7月 - 2023年10月*
腾讯AI实验室，中国

- 参与了一个旨在增强现有以英语为中心的大语言模型能力的项目，扩展其语言覆盖范围至150种自然语言和150种编程语言。这涉及对最近发布的开源LLMs在大量文本和代码语料库上进行继续预训练。升级后的LLMs在各种多语言评估基准上展示了最优性能。

访问研究员 *2023年4月 - 2023年7月*
穆罕默德·本·扎耶德人工智能大学，阿联酋

- 参与两个研究项目：(1)研究大语言模型蒸馏压缩方法，成功实现在不影响其有效性的情况下显著减小模型大小；(2)对大语言模型和人类评估者在评判机器生成文本中的偏见进行全面评估。

研究实习生 *2020年7月 - 2021年7月*
华为诺亚方舟实验室，中国

- 参与两个项目：(1)实现多数据集分布的动态平衡技术，以优化多语言和多领域机器翻译系统的训练；(2)专注于使用大量平行语料库预训练自回归和非自回归多语言机器翻译系统。

研究工程师 *2018年8月 - 2019年8月*
京东AI研究院，中国

- 开发了在线购物系统对话AI的初始版本，包括创建意图分类模型、粗粒度答案搜索引擎和细粒度排名模型。

部分发表论文

- **Minghao Wu***, Weixuan Wang*, Barry Haddow, and Alexandra Birch. "*HBO: Hierarchical Balancing Optimization for Fine-Tuning Large Language Models.*" 2025.
- **Minghao Wu***, Weixuan Wang*, Barry Haddow, and Alexandra Birch. "*ExpertSteer: Intervening in LLMs through Expert Knowledge.*" 2025.
- **Minghao Wu**, Thuy-Trang Vu, Lizhen Qu, and Gholamreza Haffari. "*The Best of Both Worlds: Bridging Quality and Diversity in Data Selection with Bipartite Graph .*" In Proceedings of the 42nd International Conference on Machine Learning (ICML). 2025. Proceedings of Machine Learning Research.
- **Minghao Wu**, Jiahao Xu, Yulin Yuan, Gholamreza Haffari, Longyue Wang, Weihua Luo, and Kaifu Zhang. "*(Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts.*" Transactions of the Association for Computational Linguistics (TACL). 2025. MIT Press.
- **Minghao Wu**, Weixuan Wang, Sinuo Liu, Huifeng Yin, Xintong Wang, Yu Zhao, Chenyang Lyu, Longyue Wang, Weihua Luo, and Kaifu Zhang. "*The Bitter Lesson Learned from 2,000+ Multilingual Benchmarks.*" 2025.
- **Minghao Wu***, Weixuan Wang*, Barry Haddow, and Alexandra Birch. "*Demystifying Multilingual Chain-of-Thought in Process Reward Modeling.*" 2025.
- **Minghao Wu***, Weixuan Wang*, Barry Haddow, and Alexandra Birch. "*Bridging the Language Gaps in Large Language Models with Inference-Time Cross-Lingual Intervention.*" In Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (ACL). 2025. Association for Computational Linguistics.
- **Minghao Wu**, and Alham Fikri Aji. "*Style Over Substance: Evaluation Biases for Large Language Models.*" In Proceedings of the 31th International Conference on Computational Linguistics (COLING). 2025. International Committee on Computational Linguistics.
- **Minghao Wu**, Thuy-Trang Vu, Lizhen Qu, and Gholamreza Haffari. "*Mixture-of-Skills: Learning to Optimize Data Usage for Fine-Tuning Large Language Models.*" In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2024. Association for Computational Linguistics.
- **Minghao Wu**, Jiahao Xu, and Longyue Wang. "*TransAgents: Build Your Translation Company with Language Agents.*" In Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2024. Association for Computational Linguistics.
- **Minghao Wu**, Abdul Waheed, Chiyu Zhang, Muhammad Abdul-Mageed, and Alham Fikri Aji. "*LaMini-LM: A Diverse Herd of Distilled Models from Large-Scale Instructions.*" In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2024. Association for Computational Linguistics.
- **Minghao Wu**, Yufei Wang, George Foster, Lizhen Qu, and Gholamreza Haffari. "*Importance-Aware Data Augmentation for Document-Level Neural Machine Translation.*" In Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2024. Association for Computational Linguistics.
- **Minghao Wu**, Thuy-Trang Vu, Lizhen Qu, George Foster, and Gholamreza Haffari. "*Adapting Large Language Models for Document-Level Machine Translation.*" 2024.
- **Minghao Wu**, George Foster, Lizhen Qu, and Gholamreza Haffari. "*Document Flattening: Beyond Concatenating Context for Document-Level Neural Machine Translation.*" In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL). 2023. Association for Computational Linguistics.

- **Minghao Wu**, Yitong Li, Meng Zhang, Liangyou Li, Gholamreza Haffari, and Qun Liu. "*Uncertainty-Aware Balancing for Multilingual and Multi-Domain Neural Machine Translation Training*." In Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2021. Association for Computational Linguistics.
- **Minghao Wu**, Fei Liu, and Trevor Cohn. "*Evaluating the Utility of Hand-crafted Features in Sequence Labelling*." In Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2018. Association for Computational Linguistics.