

MINGHAO WU

minghao.wu@monash.edu

+61 452 231 543

HIGHLIGHTS

- 8+ years of experience in research and development of machine learning, natural language processing (NLP), and generative AI.
- Strong publication record in top-tier conferences, such as ACL, EMNLP, COLING, and EACL.
- Programming Languages: Python, Java, and R.

EDUCATION

Monash University

Dec. 2021 - Jul. 2025 (expected)

Doctor of Philosophy in Computer Science

Advisors: Gholamreza Haffari, George Foster

Research Interests: Deep Learning, Natural Language Processing, Machine Translation, Multilinguality, Large Language Models, Efficiency

The University of Melbourne

Mar. 2016 - Jul. 2018

Master of Information Technology in Computing

Advisors: Trevor Cohn

The University of Sydney

Mar. 2013 - Mar. 2016

Bachelor of Science in Information Systems

EXPERIENCE

Research Intern

Jul. 2023 - Oct. 2023

Tencent AI Lab, China

- Worked on a project aimed at enhancing the capabilities of existing English-centric large language models (LLMs) by extending their linguistic coverage to include 150 natural languages and 150 programming languages. This involved further pretraining of recently released open-source LLMs on a vast collection of text and code corpora. The upgraded LLMs demonstrated the state-of-the-art performance across various multilingual evaluation benchmarks.

Visiting Researcher

Apr. 2023 - Jul. 2023

Mohamed bin Zayed University of Artificial Intelligence, UAE

- Contributed to three research projects: (1) Investigated methods for the strategic compression of large generative models, successfully achieving significant reductions in model size without compromising their effectiveness; (2) Carried out a thorough assessment of biases present in large language models (LLMs) and human evaluators in judging machine-generated text; (3) Developed efficient techniques for distilling multilingual knowledge from large generative models into more compact versions.

Research Intern

Jul. 2020 - Jul. 2021

Huawei Noah's Ark Lab, China

- Participated in two projects: (1) Implemented dynamic balancing techniques for the distribution of multiple datasets to optimize the training of multilingual and multi-domain machine translation systems; (2) Focused on pretraining both autoregressive and non-autoregressive multilingual machine translation systems using extensive parallel corpora.

- Developed the initial version of a conversational AI for an online shopping system, which involved creating an intent classification model, a coarse-grained answer search engine, and a fine-grained ranking model.

SELECTED PUBLICATIONS

- **Wu, Minghao**, Yulin Yuan, Gholamreza Haffari, and Longyue Wang. “(Perhaps) Beyond Human Translation: Harnessing Multi-Agent Collaboration for Translating Ultra-Long Literary Texts.” 2024.
- Chiyu Zhang, Yifei Sun, **Minghao Wu**, Jun Chen, Jie Lei, Muhammad Abdul-Mageed, Rong Jin, Angli Liu, Ji Zhu, Sem Park, Ning Yao, and Bo Long. “EmbSum: Leveraging the Summarization Capabilities of Large Language Models for Content-Based Recommendations.” 2024.
- **Wu, Minghao**, 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. ACL.
- **Wu, Minghao**, 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. ACL.
- **Wu, Minghao**, Thuy-Trang Vu, Lizhen Qu, George Foster, and Gholamreza Haffari. “Adapting Large Language Models for Document-Level Machine Translation.” 2024.
- **Wu, Minghao**, and Alham Fikri Aji. “Style Over Substance: Evaluation Biases for Large Language Models.” 2023.
- **Wu, Minghao**, 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. ACL.
- **Wu, Minghao**, 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. ACL.