Nannan (Amber) Huang

PhD Candidate in Natural Language Processing | Fairness and Bias Mitigation Researcher

+61 484 000 531 | amberhuang0701@gmail.com | https://www.nannanhuang.com

---

Research Focus

PhD researcher specialising in fairness and bias mitigation in AI systems with expertise in developing responsible AI solutions. Research focuses on addressing equity concerns in automated systems, with methodologies applicable to agentic AI, multi-agent systems where fairness across diverse populations is critical.

Key Skills Summary

- Specialised in natural language processing, artificial intelligence, machine learning, deep learning, and statistical modelling
- Proven expertise in curriculum design and delivery across undergraduate and postgraduate levels, including online and hybrid learning environments
- Excellence in creative thinking, problem-solving, and design thinking with demonstrated innovation in educational approaches
- Outstanding stakeholder management skills with extensive experience working with external networks, industry partners, and government bodies
- Strong leadership capabilities with proven success in team leadership, collaboration, and fostering inclusive environments for diverse student populations

---

Research Experience

PhD Computer Science – Natural Language Processing | *Mar 2022 - Mar 2026 (Expected)* | **RMIT University**
Thesis: Bias mitigation and fairness evaluation in language models for summarisation
https://doi.org/10.25439/rmt.31053415

Research Intern | *Mar 2025 – Jul 2025* | **National Institute of Informatics (NII), Japan**
Project: Detecting and Mitigating Social Bias in Text Generation Using LLMs.

DAAD Postdoc-NeT Fellowship – Natural Language Processing | Jun 2025 | German Academic Exchange Service (DAAD) Selected for a competitive international networking program connecting early-career researchers with leading AI institutions in Germany.

Teaching & Academic Service

Teaching Assistant – Foundation of AI for STEM, Machine Learning | *Jul 2023 – Current* | **RMIT University**
Teaching cohorts of 1000+ diverse STEM students in neural networks and deep learning applications; contributing to curriculum development for practical AI exercises and assessment design

Teaching Assistant – Introduction to Python, Data Wrangling | *Jan 2022 – Current* | **Monash University**
Supporting 150+ students per semester in programming and data analysis techniques; developing tutorial materials and contributing to assessment rubrics for undergraduate data science courses

Honours Student Supervisor | *2025 – Current* **| RMIT University**
Supervising honours student research on LLM alignment using RLHF (Reinforcement Learning from Human Feedback), focusing on data efficiency and reliability in agent alignment methodologies

---

Education

PhD of Computer Science | *Mar 2022 – Dec 2025 (Expected)* | **RMIT University**
Master of Data Science | *July 2020 – Nov 2021* | **RMIT University**
Master of Data Analytics | *Mar 2019 – July 2020* | **RMIT University**
Bachelor of Commerce (Accounting & Finance) | *Feb 2014 - Nov 2016* | **University of Melbourne**

Technical Expertise

NLP & AI Fairness: Transformer models, PyTorch, bias evaluation frameworks, large language model assessment, fairness methodologies for diverse populations, multi-agent LLM systems, agent orchestration and coordination, LLM prompting and tool use, safety evaluation methodologies for AI
Data Science: Scikit-learn, statistical analysis, experimental design, large-scale dataset processing (Pandas, NumPy, SciPy)
Research Skills: Evaluation framework development, interdisciplinary collaboration, bias detection and mitigation

---

Key Publications

Huang, N., Maab, I., & Yamagishi, J. (Under Review). 'When Bigger Isn't Better: A Comprehensive Fairness Evaluation of Political Bias in Multi-News Summarisation'
Develops fairness evaluation methodologies adaptable for NLP systems processing diverse content

Huang, N., Fayek, H., & Zhang, X. J. (2025). 'REFER: Mitigating Bias in Opinion Summarisation via Frequency Framed Prompting' *NewSumm Workshop at EMNLP 2025*
Novel approaches for reducing demographic bias in summarisation systems

Huang, N., Fayek, H., & Zhang, X. J. (2025). 'Less Is More? Examining Fairness in Pruned Large Language Models for Summarising Opinions' *EMNLP 2025*
Addresses computational efficiency while maintaining equitable performance for resource-constrained applications

Huang, N., Fayek, H., & Zhang, X. J. (2024). 'Bias in Opinion Summarisation from Pre-training to Adaptation: A Case Study in Political Bias.' *EACL 2024*
Establishes bias detection methodologies and evaluates across adaptation techniques

Huang, N., Tian, L., Fayek, H., & Zhang, X. J. (2023). 'Examining Bias in Opinion Summarisation through the Perspective of Opinion Diversity.' *WASSA Workshop at ACL 2023*
Investigates fairness through diverse viewpoint representation in automated summarisation

Huang, N., & Zhang, X. (2021). 'Evaluation of Review Summaries via Question-Answering.' *ALTA 2021*

Develops question-answering based assessment techniques for summarisation quality and completeness

Additional Collaborative Publications:

Huang, N., Zhang, X. J., & Du, J. T. (2025). 'AI for human and misinformation interactions: A case of social media.' In D. Wu & S. Liang (Eds.), Human-AI interaction and collaboration. Cambridge University Press

Tian, L., Huang, N., & Zhang, X. (2024). 'Large language model cascades and persona-based in-context learning for multilingual sexism detection.', International Conference of the Cross-Language Evaluation Forum for European Languages 2024

Tian, L., Huang, N., & Zhang, X. (2023). 'Efficient Multilingual Sexism Detection via Large Language Model Cascades.', CLEF 2023