WEIJUN LI

Sydney, Australia | +61-0472671546 | weijun.li1@hdr.mq.edu.au | weijun-l.github.io | Google Scholar

Education

- **Ph.D. in Computer Science** Macquarie University Sydney, Australia Feb 2025 – Feb 2028 (expected)
  - Supervisors: Prof. Mark Dras and Dr. Qiongkai Xu
- **Master of Research in Computing** Macquarie University Sydney, Australia Jan 2024 – Dec 2024
  - Overall Grade: 93.0% (High Distinction)
  - Supervisors: Prof. Mark Dras and Dr. Qiongkai Xu
- **Master of Information Technology** The University of Melbourne Melbourne, Australia Feb 2022 – Dec 2023
  - Overall Grade: 82.3% (First Class Honours)
  - Research Thesis Grade: 90.0% (Supervisor: Dr. Qiongkai Xu)
- **Master of Control Engineering** Chongqing University Chongqing, China Sep 2013 – Jun 2016
- **Bachelor of Mechanical Engineering** Chongqing University Chongqing, China Sep 2009 – Jun 2013

Research Interest

I study **trustworthy machine learning**, focusing on **privacy** and **security** in language and emerging multimodal models. My work spans two directions: (1) data privacy, including leakage auditing and empirical privacy calibration; and (2) model security, especially post-training backdoor defenses. The outcomes include practical auditing and mitigation methods for real-world deployment.

Publications

Preprint ***Beyond Theoretical Bounds: Empirical Privacy Loss Calibration for Text Rewriting Under Local Differential Privacy.***
Weijun Li, Arnaud Grivet Sébert, Qiongkai Xu, Annabelle McIver, and Mark Dras.
arXiv preprint, 2026. [Link]

ICLR 2026 ***Defending against Backdoor Attacks via Module Switching.***
Weijun Li, Ansh Arora, Xuanli He, Mark Dras, and Qiongkai Xu.
In *The Fourteenth International Conference on Learning Representations*. [Link]

EMNLP 2025 ***Cut the Deadwood Out: Backdoor Purification via Guided Module Substitution.***
Yao Tong*, **Weijun Li***, Xuanli He, Haolan Zhan, and Qiongkai Xu.
In *Findings of the Association for Computational Linguistics: EMNLP 2025*. [Link]

EMNLP 2024 ***Seeing the Forest through the Trees: Data Leakage from Partial Transformer Gradients.***
Weijun Li, Qiongkai Xu, and Mark Dras.
In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*. [Link]

* Equal contribution.

Academic Leadership & Service

- **Organizer**, *Anti-BAD: An Anti-Backdoor Challenge for Post-Trained Large Language Models* [Link]
  - Led proposal writing and overall challenge design, including task formulation and evaluation protocol.
  - Developed and managed the competition platform on Codabench for submission and evaluation.
  - Presented the challenge at IEEE SaTML 2026 (Munich), summarizing design, results, and key findings [Link].
- **Reviewer**: ACL Rolling Review (ARR), 2025–2026 (for ACL, EMNLP, and others); NeurIPS 2026 (Invited Reviewer)

Research Projects

- **Ph.D. Research: Addressing Data-Driven Vulnerabilities in Language Models** Macquarie University Feb 2025 – Present
  - Audit privacy risks arising from data interactions in language models, and develop empirical calibration methods for local differential privacy (*Preprint*).
  - Develop backdoor defense methods for post-trained language models, including module switching and guided module substitution (*ICLR 2026; EMNLP 2025*).
- **M.Res. Thesis: Data Reconstruction Attacks in Distributed Learning** Macquarie University Jan 2024 – Dec 2024
  - Showing that gradients from only 0.54% of model parameters can leak private text data, substantially strengthening gradient inversion attacks (*EMNLP 2024*).

Honors & Recognition

- Executive Dean's Commendation for Academic Excellence, Master of Research, Macquarie University 2025
- HDR Research Rising Star Award, School of Computing, Macquarie University 2024
- Dean's Honours List, Faculty of Engineering and Information Technology, The University of Melbourne 2023

Professional Experience

- **Teaching Associate** 2025
  COMP8420 Advanced Natural Language Processing, Macquarie University Sydney, Australia
- **Automotive R&D Engineer** Jul 2016 – Dec 2021
  Dongfeng-Nissan Passenger Vehicle Company Guangzhou, China