Lu Xinyang

xinyang.lu@u.nus.edu | +65 90872041

EDUCATION

Doctor of Philosophy (School of Computing)   *Aug 2022 – present*

National University of Singapore

Research Interests: Large Language Model, Machine Unlearning

Bachelor of Engineering (Electrical & Electronics Engineering)   *Aug 2018 – Jun 2022*

Nanyang Technological University

Cumulative Grade Point Average: 4.8/5

Relevant Courses: Engineering Mathematics, Data Structures & Algorithms, Computer Communications, Artificial Intelligence & Data Mining, Digital Signal Processing

Exchange Experience:

- École Polytechnique Fédérale de Lausanne, Switzerland   *Feb 2020 – Apr 2020*
- University of California, Berkeley, USA   *May 2019 – Jun 2019*

PUBLICATIONS

- Lu, X.*, Niu, X.*, Lau, G.*, Bui, N., Sim, R. H. L., Himawan, J. R., Wen, F., Foo, C. S., Ng, S., & Low, B. K. H. (2026). WaterDrum: Watermarking for Data-centric Unlearning Metric. In *Proc. ICLR*.
- Bui, N.*, Lu, X.*, Ng, S., & Low, B. K. H. (2026). On Newton's Method to Unlearn Neural Networks. In *Proc. ICLR*.
- Lu, X.*, Wang, J.*, Zhao, Z.*, Dai, Z., Foo, C. S., Ng, S. K., & Low, B. K. H. (2025). WASA: Watermark-based source attribution for large language model-generated data. *Findings of ACL*.
- Agussurja. L., Lu, X., & Low, B. K. H. (2024). Global-to-Local Support Spectrums for Language Model Explainability. *arXiv preprint arXiv:2408.05976*.
- Xiong, M., Hu, Z., Lu, X., LI, Y., Fu, J., He, J., & Hooi, B. (2024). Can LLMs Express Their Uncertainty? An Empirical Evaluation of Confidence Elicitation in LLMs. In *Proc. ICLR*.
- Lu, X., Fan, F. X., & Wang, T. (2023). Action and Trajectory Planning for Urban Autonomous Driving with Hierarchical Reinforcement Learning. In *ICML Workshop on New Frontiers in Learning, Control, and Dynamical Systems*.
- Farah, J. C., Spaenlehauer, B., Lu, X., Ingram, S., & Gillet, D. (2022). An exploratory study of reactions to bot comments on github. In *Proc. International Workshop on Bots in Software Engineering*.

WORKING EXPERIENCE

Research Assistant   *Aug 2022 – Jun 2025*

Institute of Data Science, National University of Singapore, Singapore

- Conduct research on Collaborative Machine Learning related topics
- Participate in several papers, which have been submitted and under review in conferences.

Research Intern   *Jan 2021 – Jun 2021*

Agency for Science, Technology and Research, Singapore

- Conducted research of the application of reinforcement learning on autonomous vehicles

- Developed a DDPG based model with intention and trajectory planners to deal with dynamic and completed urban driving scenarios

Web Development Intern   *Dec 2019 – Jan 2020*
Tong Cheng International Travel Agency Co., Ltd., China
- Internship in Java Web design and development in the travel agency cooperation
- Analyzed the application of MVC framework and servlet on web development
- Took part in the development of the front-end and background services of the travel agency website

SKILLS

Languages: Chinese (Native), English (Fluent)
Interests: Large Language Models, Machine Unlearning, Data-centric AI, Watermarking