Jia Peng Lim

Email: jiapeng.lim.2021@phdcs.smu.edu.sg | Socials: jararap.github.io

Singaporean Citizen, Singapore

Summary

Natural Language Processing Researcher with 4+ years of experience conducting and shipping AI research, specializing in Large Language Models and representation learning. I design rigorous experiments and develop scalable prototypes in Python, C++, Java, and PyTorch, with first-author publications at top-tier venues including ACL, EMNLP, and NeurIPS. I am adept at rapidly ramping up in new research areas and translating unfamiliar problem domains into well-defined, empirical research programs.

Education

Ph.D. Computer Science, Singapore Management University   08/2021 – 05/2026 (Expected)

B.Sc. Information Systems, Singapore Management University   08/2017 – 05/2021

    cGPA: 4.00/4.00, Summa Cum Laude

Publications

A Partition Cover Approach to Tokenization, *NeurIPS 2025* [link]

    Jia Peng Lim, Shawn Tan, Davin Choo, and Hady W. Lauw

Interpreting Topic Models in Byte-Pair Encoding Space, *COLING 2025* [link]

    Jia Peng Lim and Hady W. Lauw

Aligning Human and Computational Coherence Evaluations, *CL Journal 2024* [link].

    Jia Peng Lim and Hady W. Lauw

Disentangling Transformer Language Models as Superposed Topic Models, *EMNLP 2023 (Main)* [link]

    Jia Peng Lim and Hady W. Lauw

Large-Scale Correlation Analysis of Automated Metrics for Topic Models, *ACL 2023 (Main)* [link]

    Jia Peng Lim and Hady W. Lauw

Towards Reinterpreting Neural Topic Models via Composite Activations, *EMNLP 2022 (Main)* [link]

    Jia Peng Lim and Hady W. Lauw

Awards

DSO-AISG Incentive Award (2026)

AISG Research PhD Fellowship Award (2023 – 2025)

Singapore Data Science Consortium Dissertation Research Fellowship (2023)

SMU Presidential Doctoral Fellowship (2023 – 2025)

Ph.D. Dean's List (2024)

PayPal Fintech Scholarship (2018 – 2021)

Undergraduate Dean's List (2018, 2019, 2020)

Top First Year Information Systems Student (2018)

Work Experience

Research Scientist (Intern), DSO National Laboratories, Singapore 01/2026 – Current

Investigating and developing methods to improve domain-specific explainability in foundation models for multi-modal reasoning and to enhance multilingual reasoning in low-resource languages. This work is supported by the DSO–AISG Incentive Award.

AI Robotics Intern, Bosch Digital Innovation, Singapore 08/2025 – 12/2025

Designed and implemented an unsupervised, automated pipeline for multi-modal knowledge graph construction, leveraging classical and deep-learning computer vision (OpenCV, PyTorch) to extract spatial features from images and integrate them into downstream robotic perception systems for manufacturing automation.

Data Engineering Intern, Lucence, Singapore 05/2020 – 08/2020

Built production-grade, serverless ETL and data ingestion pipelines on AWS using Python, automating large-scale web crawling and dataset curation to automate research workflows in the biotechnology domain.

Software Engineering Intern, PayPal, Singapore 05/2019 – 08/2019

Translated academic research on Conflict-free Replicated Data Types (CRDTs) into a production Java library, enabling active-active distributed data consistency and fault tolerance for payment systems requiring high availability and low-latency data access.

Community Service

Senior Teaching Assistant for Undergraduate Courses 08/2025 – 11/2025

1. Smart-City Operations Research
2. Managing Customer Relations in Asia

Instructor for IMDA-Google AI Accelerator 06/2025

Served as one of the instructors in a team that taught high school students about AI (Computer Vision) using Python and PyTorch during a one-week intensive bootcamp in the Summer of 2025. Using a novel dataset provided by sponsors, I also designed a final project problem, which the students worked on for a week before their final presentation.

Conference Organizing Committee

Served as the aide-de-camp to the programme chair of ACM Web Search and Data Mining (WSDM) 2023 and TheWebConf (WWW) 2024.