

Jia Peng Lim

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

Singaporean Citizen, Singapore

Summary

Research Experience of 4+ years, ideating and executing Artificial Intelligence (AI) research projects, including research on Large Language Models. Using Python, C++, Java and Machine Learning frameworks (e.g. PyTorch), I prototype and pursue research objectives, publishing first-author research papers in Top AI conferences such as ACL, EMNLP, and NeurIPS.

Education

<i>Ph.D. Computer Science, Singapore Management University</i>	08/2021 – 05/2026 (Expected)
<i>B.Sc. Information Systems, Singapore Management University</i>	08/2017 – 05/2021

Publications

- A Partition Cover Approach to Tokenization, *NeurIPS 2025 (to appear)* [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

- 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

AI Robotics Intern, Bosch Digital Innovation, Singapore

08/2025 – 12/2025

Proposed and developed an automated approach to build a multi-modal knowledge graph, extracting information from images using traditional and deep learning/GenAI computer vision techniques (OpenCV, PyTorch), in the context of augmenting spatial information fed to fixed-armed robots for manufacturing automation.

Data Engineering Intern, Lucence, Singapore

05/2020 – 08/2020

Build serverless Extra, Transform, Load (ETL) pipelines and applications hosted on Amazon Web Services, to curate and serve customized data requirements for organizational research needs in biotechnology space.

Software Engineering Intern, PayPal, Singapore

05/2019 – 08/2019

Research, designed and implemented a Java library on Conflict-free Replicated Data Types. The subsequent code went into production, assisting in the organization's goal of full active-active redundancy.

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.