LEE JUNYOUNG

Curriculum Vitae WORK & RESEARCH EXPERIENCE

✓ junyounglee.k@gmail.comin jyleek♠ jylee-k♦ (+65) 9451 5623

Home Team Science and Technology Agency Enterprise Group - xData

Singapore

Data Scientist (2024.07 - present)

- Designing proof-of-concept implementation of community detection in customized multilayer graph networks
- Developing feature extraction methods for graph neural network training
- Technical lead for deployment of a containerized application for customized graph analytics

Infocomm Media Development Authority BizTech Group

Singapore

Al Developer Intern (2024.01 - 2024.05)

- ➤ Benchmarking evaluation system for generative AI (2024.01 2024.05)
- Implemented benchmarking pipeline comparing in-house finetuned LLM to open-source LLMs
- Developed synthetic QA dataset generation framework for LLM evaluation

Tokyo Institute of Technology Okazaki Laboratory

Tokyo, Japan

Visiting Research Student (2023.07 - 2024.01)

- Subword tokenization and regularization for machine translation (partially remote, 2023.07 present)
- Investigated Korean jamo-level byte-pair encoding for English to Korean neural machine translation
- Developed a constrained generation method for Korean jamo sequence based on tensor masking
- Investigating subword tokenization for Jejueo, a critically endangered Koreanic language

Accredify SGInnovate Summation Programme

Singapore

Data Science Intern (2023.05 - 2023.11)

- ➤ Retrieval-augmented generation (2023.05 2023.07)
- Developed retrieval augmented generation-based solution using LLMs for company database querying and customer support chatbot deployment on AWS
- ➤ Management dashboard development (2023.08 2023.11)
- Designed data ingestion pipeline and framework for internal management dashboard on AWS and Tableau

École Polytechnique Fédérale de Lausanne Digital and Cognitive Musicology Lab

Lausanne, Switzerland

Research Intern (2022.05 - 2022.08)

- ➤ Contextual chord embedding (2022.05 2022.08)
- Conducted experiments to develop contextual chord embeddings with symbolic harmonic features via multitask transfer learning

Nanyang Technological University

Singapore

Undergraduate Research Student (2020.12 - 2024.06) • Research Affiliate (2024.07 - present)

- · Developed phrase feature extraction for graph construction and classification from question text
- Designing novel graph convolutional network architecture for multi-graph embedding aggregation
- ➤ Teamwork dimensions classification (2022.09 2023.07)
- Finetuned pretrained language model on classroom discussion data for teamwork dimensions for learning analytics system development

EDUCATION

BEng (Hons) in Electrical and Electronic Engineering

Singapore

Nanyang Technological University • Honours (Highest Distinction) • CGPA 4.89/5.00

2020.08 - 2024.06

EEE Chair Award • EEE Excellence Award • Dean's Lists AY20/21, 23/24 Awards:

DSP System Design • Computational Intelligence • Pattern Recognition & Deep Learning Coursework: Semester Exchange: National University of Singapore, Singapore (2022) • University of Waterloo, Canada (2023)

SCHOLARSHIP

JASSO Scholarship (2023)

EPFL Scholarship of Excellence (2022)

ThinkSwiss Research Scholarship Asia-Pacific (2022)

Nanyang Scholarship + CN Yang Scholars Programme (2020 - 2024)

Japan Student Services Organization École Polytechnique Fédérale de Lausanne **Embassy of Switzerland in Singapore** Nanyang Technological University, Singapore

PROFESSIONAL SERVICE

Reviewer

International Conference for Artificial Intelligence in Education (AIED; 2024)

International Joint Conference on Neural Networks (IJCNN; 2025)

PUBLICATIONS

Preprint / Under Review

* indicates equal contribution

J. Lee*, M. Cognetta*, S. Moon, and N. Okazaki

Submitted 2025.02

Jamo-Level Subword Tokenization in Korean Machine Translation

Under review for The Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT 2025)

J. Lee, N. Dixit*, K. Chakrabati*, and S. Supraja

Submitted 2025.01

Word and Phrase Features in Graph Convolutional Network for Automatic Question Classification

Under review for INNS International Joint Conference on Neural Networks (IJCNN) 2025

Journal

K. N. Watcharasupat, J. Lee, and A. Lerch

2022.02

Latte: Cross-framework Python package for evaluation of latent-based generative models 🖺 😱

Software Impacts, vol. 11, art. no. 100222

Conference (peer reviewed)

J. Lee. and E. Koh

2023.07

Teamwork Dimensions Classification Using BERT

Proceedings of the 24th International Conference on Artificial Intelligence in Education (AIED 2023) Posters and Late Breaking Results, pp. 254-259

Extended Abstracts (peer reviewed)

Y.-N. Hung, K. N. Watcharasupat, C.-W. Wu, I. Orife, K. Li, P. Seshadri, and J. Lee

2021.11

AVASpeech-SMAD: A Strongly Labelled Speech and Music Activity Detection Dataset with Label Co-Occurrence

Late-Breaking Demo Session of the 22nd International Society for Music Information Retrieval Conference (ISMIR)

SKILLS

Programming Python MATLAB C

PyTorch + Lightning TensorFlow + Keras Pandas scikit-learn Networkx ML/AI

AWS GCP Azure Cloud

Prototyping Fusion 360 Arduino Raspberry Pi ESP32

Performance Trumpet Flugelhorn Cornet Euphonium

Languages English (bilingual) Korean (native) Japanese (basic)