

LEE JUNYOUNG

Curriculum Vitae

WORK & RESEARCH EXPERIENCE

✉ junyounglee.k@gmail.com
in jyleek
🗣 jylee-k
☎ (+65) 9451 5623

ZenLow

📍 Remote

Co-Founder (2026.04 – present)

✓ Co-CTO (2026.04 – present)

- Building and managing product UI/UX for merchant, buyer, and admin interfaces; SEO management

Home Team Science and Technology Agency Enterprise Group - xData

📍 Singapore

Data Scientist - S&T Associate (2024.07 – present)

✓ Graph network analysis for risk sense-making (2024.07 – present)

- Designing proof-of-concept implementation of node classification in customized multi-layer graph networks
- Developing text feature extraction methods for graph-based entity resolution
- Technical lead for deployment of a containerized application for customized graph analytics

✓ Ops-tuned LLM (2025.02 – 2026.05)

- Built framework for finetuning small LMs to improve alignment with human annotators in LLM-as-a-Judge tasks
- Built and benchmarking RAG interface on edge devices using quantized models
- Optimized synthetic data generation for LLM finetuning in Singapore's Home Team context
- Evaluated small LM pre-training and finetuning framework with TinyLlama & SeaLLM researchers from SUTD

Infocomm Media Development Authority BizTech Group

📍 Singapore

AI Developer Intern (2024.01 – 2024.05)

✓ Benchmarking evaluation system for generative AI (2024.01 – 2024.05)

- Implemented pipeline comparing in-house finetuned LLM to open-source LLMs for benchmarks of bias, cause-and-effect, contextual parametric knowledge conflicts and toxicity
- Developed synthetic QA dataset generation framework for LLM evaluation and finetuning

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 – 2025.05)

- 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
- Investigated subword tokenization for Jejueo, a critically endangered Koreanic language, for low-resource machine translation

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 for KPI monitoring using 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 with LSTM and Transformer-based models

EDUCATION

MComp in Artificial Intelligence

National University of Singapore • Dissertation Track • CGPA 5.00/5.00

Singapore
2025.08 – 2027.12 (expected)

BEng (Hons) in Electrical and Electronic Engineering

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

Singapore
2020.08 – 2024.06

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

Semester Exchange: National University of Singapore, Singapore (2022) • University of Waterloo, Canada (2023)

SCHOLARSHIP

JASSO Scholarship (2023)

Japan Student Services Organization

EPFL Scholarship of Excellence (2022)

École Polytechnique Fédérale de Lausanne

ThinkSwiss Research Scholarship Asia-Pacific (2022)

Embassy of Switzerland in Singapore

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

Nanyang Technological University, Singapore

PROFESSIONAL SERVICE

Reviewer/PC Member

International Conference for Artificial Intelligence in Education (AIED; 2024 - 2026)

International Joint Conference on Neural Networks (IJCNN; 2025 - 2026)

SELECTED PUBLICATIONS

Peer Reviewed

* indicates primary authorship

M. Cогnetta, D. Pohl, J. Lee, and N. Okazaki

2025.07

Pitfalls, Subtleties, and Techniques in Automata-Based Subword-Level Constrained Generation

Proceedings of the ICML 2025 Tokenization Workshop (TokShop)

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

2025.07

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

International Joint Conference on Neural Networks (IJCNN 2025) Poster Track

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

2025.05

Jamo-Level Subword Tokenization in Korean Machine Translation

Proceedings of the Eighth Workshop on Technologies for Machine Translation of Low-Resource Languages (LoResMT) @ NAACL 2025

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

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

Preprint / Under Review

J. Wang and J. Lee*

2026.04

More Human, More Efficient: Aligning Annotations with Quantized SLMs

arXiv:2604.00586

D. Pohl, M. Cогnetta, J. Lee, and N. Okazaki

2025.10

Decoding-Free Sampling Strategies for LLM Marginalization

arXiv:2510.20208

SKILL

Programming

Python MATLAB C SQL

ML/AI

PyTorch + Lightning TensorFlow + Keras Pandas scikit-learn Networkx

Cloud

AWS GCP Azure

Prototyping

Fusion 360 Arduino Raspberry Pi ESP32

Dashboard

Tableau Power BI

Languages

English (fluent) Korean (native) Japanese (basic)