# Jaewook (Jake) Lee

LLM Post-Training & Evaluation | Applied NLP
Amherst, MA | jaewooklee@umass.edu | LinkedIn | Google Scholar | jwlee.info

## Summary

PhD candidate in Computer Science at the University of Massachusetts Amherst specializing in LLM post-training, evaluation, agentic RAG, and conversational AI. Builds and evaluates language-model systems for chat-based interaction, personalization, user simulation, retrieval, and human-AI collaboration, using methods such as reinforcement learning, tool use, activation steering, and human-centered evaluation. Industry experience includes multimodal search at Apple and evaluation methodology for production multi-agent security agents at AWS. Author of 10+ publications at venues including ACL, EMNLP, NAACL, and CHI.

## Technical Skills

**Core**: LLM Post-Training, Reinforcement Learning, SFT, Preference Optimization, LLM Evaluation, Agentic RAG, Personalization, Dialogue Systems, User Simulation
**Systems**: Chat-Based AI Systems, Retrieval Systems, Tool Use, Multi-Agent Systems, Large Language Models, Vision-Language Models, vLLM, RAG Accelerators, AI Accelerators
**Evaluation & Alignment**: Human Evaluation, Benchmark Design, Evaluator Design, Reward Modeling
**Engineering**: Python, PyTorch, Hugging Face, iOS, React, Go, JAX, Ray, Kubernetes

## Experience

**Apple** Jun. 2026 – Present
*Machine Learning Engineer Intern* *Santa Clara, CA*
- Contributing to multimodal search systems for Apple's Answers, Knowledge, and Information team; additional technical details limited by confidentiality.

**Amazon Web Services** May 2025 – Aug. 2025
*Applied Scientist Intern* *Seattle, WA*
- Designed the evaluation methodology for a production multi-agent AppSec framework, defining task rubrics and assessments for design reviews, pull-request code reviews, and context-aware penetration testing.

**Eedi** Jun. 2023 – Aug. 2023
*Machine Learning Research Intern* *Remote, London, UK*
- Prototyped AnSearch, an LLM-based diagnostic-generation system that combines model outputs with domain expertise to create targeted assessments for identifying student misconceptions.

## Selected Research

**Agentic RAG, Tool Use & Retrieval Systems**

- Built an agentic RAG policy that trains LLMs with reinforcement learning to choose retrieval granularity and coordinate keyword/semantic search tools (Submitted to EMNLP 2026).
- Evaluated latency–accuracy trade-offs in software–hardware co-designed RAG, analyzing how approximate vs. exact nearest-neighbor retrieval affects time-to-first-token and downstream answer quality (Manuscript in preparation).

**Post-Training, Dialogue Systems & Model Behavior**

- Developing SAE-based interpretability methods to identify latent dialogue traits and steer LLM behavior toward interpretable user characteristics (Ongoing project).
- Built an evaluation framework for LLM-simulated users in tutoring dialogues, measuring linguistic, behavioral, and cognitive realism and showing that surface-level plausibility can mask failures in learning dynamics (ACL 2026).
- Developed a preference-optimization method for deriving tutor-persona steering vectors from human tutor-student dialogues, enabling interpretable control of LLM dialogue behavior without explicit persona prompting (ACL 2026 BEA Workshop).
- Developed an Expectation Maximization-based personalization framework that infers community-level and user-specific latent preferences, enabling LLM generation under cold-start conditions (EMNLP 2025).

**Applied NLP, Personalization & Human-Centered Evaluation**

- Built an iOS prototype for adaptive LLM generation that personalizes learning content to individual user preferences through interactive feedback (CHI EA 2026 ).
- Led research on phonological similarity algorithms for cross-lingual keyword discovery grounded in user language priors, supporting personalized mnemonic generation (EMNLP 2025).
- Designed evaluation pipelines combining psycholinguistic metrics with human assessment to measure memorability and creative quality in LLM-generated mnemonics (EMNLP 2024).

## Education

**University of Massachusetts Amherst** — Amherst, MA

*Ph.D. in Computer Science* — *Sep. 2022 – Present*

- Advisor: Dr. Andrew Lan
- Research: LLM Post-Training and Evaluation, Personalization, Human-in-the-Loop AI
- Mentored student-led ML/NLP research projects, including collaborations with Spotify and Adobe Research
- Passed qualifying exam with distinction (May 2025); M.S. earned Feb. 2026

**Korea University** — Seoul, South Korea

*M.E. & B.E. in Electrical and Computer Engineering* — *Mar. 2013 – Feb. 2022*

- M.E. (Jan. 2019 – Feb. 2022), Advisor: Dr. Seon Wook Kim
- Research: Compiler Optimization, Processing-in-Memory, AI Frameworks
- Developed ONNX Runtime integration and profiling/scheduling algorithms for PIM-based DNN inference with SK Hynix
- Built NPU compiler tooling for memory-trace extraction and scratchpad-aware LLVM code generation with Samsung
- B.E. (Mar. 2013 – Aug. 2019), Graduated with honors; Mandatory Military Service (21 months)

## Selected Publications

[1] Varun Gandhi*, **Jaewook Lee***, Shantanu Todmal, Franck Dernoncourt, Ryan Rossi, Zichao Wang, Andrew Lan, "GRASP: Granularity-Aware Search Policy for Agentic RAG." (Submitted to EMNLP 2026)

[2] Donghyeon Joo, **Jaewook Lee**, Sanjali Yadav, Ramyad Hadidi, Thomas Goldstein, Bahar Asgari, "RYLOTH: Accelerating RAG with Dynamically Reconfigurable Retrieval Architecture." (Manuscript in preparation)

[3] Alexander Scarlatos, **Jaewook Lee**, Simon Woodhead, Andrew Lan, "Simulated Students in Tutoring Dialogues: Substance or Illusion?" *Annual Meeting of the Association for Computational Linguistics (ACL)*, San Diego, CA, USA, 2026. [Link]

[4] **Jaewook Lee**, Alexander Scarlatos, Simon Woodhead, Andrew Lan, "Letting Tutor Personas Speak Up for LLMs: Learning Steering Vectors from Dialogue via Preference Optimization," *Workshop on Innovative Use of NLP for Building Educational Applications (BEA)*, 2026. [Link]

[5] **Jaewook Lee**, Alexander Scarlatos, Andrew Lan, "Interpretable Mnemonic Generation for Kanji Learning via Expectation-Maximization," *Proceedings of the Association for Computational Linguistics: EMNLP*, Suzhou, China, 2025. [Link]

[6] **Jaewook Lee**, Wooho Park, Jennifer Kyungeun Lee, Jonggi Hong, Yuki Yoshimura, Andrew Lan, "Personalizing Kanji Memorization: Designing Adaptive Mnemonics Based on Learner Preferences Using Large Language Models," *ACM Conference on Human Factors in Computing Systems (CHI) Extended Abstracts*, Barcelona, Spain, 2026. [Link]

[7] Sana Kang*, Myeongseok Gwon*, Su Young Kwon*, **Jaewook Lee**, Andrew Lan, Bhiksha Raj, Rita Singh, "PhoniTale: Phonologically Grounded Mnemonic Generation for Typologically Distant Language Pairs," *Proceedings of the Association for Computational Linguistics: EMNLP*, Suzhou, China, 2025. [Link]

[8] **Jaewook Lee**, Jeongah Lee, Wanyong Feng, Andrew Lan, "From Text to Visuals: Using LLMs to Generate Math Diagrams with Vector Graphics," *International Conference on Artificial Intelligence in Education*, Palermo, Italy, 2025. [Link]

[9] Alexander Scarlatos, Naiming Liu, **Jaewook Lee**, Richard Baraniuk, Andrew Lan, "Training LLM-based Tutors to Improve Student Learning Outcomes in Dialogues," *International Conference on Artificial Intelligence in Education*, Palermo, Italy, 2025. [Link]

[10] **Jaewook Lee**, Hunter McNichols, Andrew Lan, "Exploring Automated Keyword Mnemonics Generation with Large Language Models via Overgenerate-and-Rank," *Findings of the Association for Computational Linguistics: EMNLP*, Florida, USA, 2024. [Link]

[11] Wanyong Feng*, **Jaewook Lee***, Hunter McNichols*, Alexander Scarlatos*, Digory Smith, Simon Woodhead, Nancy Otero Ornelas, Andrew Lan, "Exploring Automated Distractor Generation for Math Multiple-choice Questions via Large Language Models," *Findings of the Association for Computational Linguistics: NAACL*, Mexico City, Mexico, 2024. [Link]

[12] Seok Young Kim*, **Jaewook Lee***, Yoonah Paik, Chang Hyun Kim, Won Jun Lee, Seon Wook Kim, "Optimal Model Partitioning with Low-Overhead Profiling on the PIM-based Platform for Deep Learning Inference," *ACM Transactions on Design Automation of Electronic Systems*, 29(2), 1–22. [Link]

[13] **Jaewook Lee**, Andrew Lan, "SmartPhone: Exploring Keyword Mnemonic with Auto-generated Verbal and Visual Cues," *International Conference on Artificial Intelligence in Education*, Tokyo, Japan, 2023. [Link]

## Awards

**2023** – NAEP Math Automated Scoring Challenge Grand Prize

**2022** – NeurIPS 2022 Causal Edu Competition (Task 3), 3rd Place

**2018** – iOS Application Hackathon Grand Prize, Korea University

## References

**Andrew Lan**: Associate Professor, University of Massachusetts Amherst; andrewlan@umass.edu

**Alexander Scarlatos**: Research Engineer, Duolingo; ajscarlatos@umass.edu

**Simon Woodhead**: Co-founder and CTO, Eedi; simon.woodhead@eedi.com