Daniel Hyunsoo Lee

dani.cmh.lee@gmail.com | dhl9@illinois.edu | github.com/danihyunlee | danihyunlee.github.io | Cell: (516)-287-7851

EDUCATION

University of Illinois, Urbana-Champaign (UIUC) Urbana, IL
Master of Science (Funded Thesis-Based) in Computer Science *Aug 2026 – Expected May 2028*
Advised by Professor Jiaqi Ma

Columbia University New York, NY
Bachelor of Engineering in Computer Science (Intelligent Systems Track), Minor in Economics *Aug 2018 – May 2022*

- **GPA:** 4.16/4.33 (4.20 Major GPA), SAT: 750/800 (M/V). Graduated *Summa Cum Laude*
- **Honors / Activities:** Egleston Scholar Grant (Merit-based Tuition / Research Grant); Dean's List; VP, Columbia Korean Student Association

RESEARCH EXPERIENCE

TRAIS Lab at UIUC Urbana, IL
Machine Learning Research Assistant under Jiaqi Ma *Aug 2026 – Present*

- Developing inference-time scaling and discovery-oriented post-training algorithms and memory mechanisms for LLMs and world models.

Improbable AI Lab at Massachusetts Institute of Technology Remote
Research Collaborator under Pulkit Agrawal *Apr 2025 – Present*

- Led research on synthetic pre-training methods for foundation model pre-training using neural cellular automata, designing algorithms and experiments to induce transferable biases for foundation models across language modeling, reasoning, coding, vision, and genomics
- Engineered and troubleshooted multi-node distributed pre-training and data curation workflows in PyTorch across Slurm clusters to efficiently pre-train LLMs of 3B+ active parameters for experiments and iterative development of new synthetic data algorithms
- In our paper [1], we show that pre-pre-training LLMs on synthetic data generated from neural cellular automata (NCAs) can improve language pre-training efficiency by up to 40% and perplexity by up to 6% while also improving model reasoning on GSM8K, Humaneval, and Big-Bench tasks with 10x less compute and data than pre-pretraining on natural language

PUBLICATIONS & PREPRINTS

[3] Breslow, N., Han, S., **Lee, D.**, Mishra, A., Liu, A., Khashabi, D. (2026). Convergent Emergence of In-Context Learning Across Modalities. In submission for ICLR 2027; arXiv preprint arXiv:2609.14011; https://arxiv.org/abs/2609.14011

[2] **Lee, D.***, Han, S.*, Breslow, N., Revsine, M., Khashabi, D., Agrawal, P. (2026). Pre-Pre-Training Before Pixels and Genes. **Spotlight Presentation,** *NeurIps 2026 Workshop on Linguistic Principles for Foundation Models***.**

[1] **Lee, D.***, Han, S.*, Kumar, A., Agrawal, P. (2026). Training Language Models via Neural Cellular Automata. NeurIPs 2026; arXiv preprint arXiv:2603.10055; https://hanseungwook.github.io/blog/nca-pre-pre-training/

TEACHING EXPERIENCE

University of Illinois-Urbana-Champaign Urbana, IL
Teaching Assistant for CS580: Algorithmic Game Theory *Fall 2026*

Columbia University New York, NY
(Head) Teaching Assistant *Fall 2020 - Fall 2021*

- Led recitations on complexity analysis, data structures (stacks, queues, graphs), algorithms, and implementations in Java and C++
- Managed homework and exam grading for Columbia's largest undergraduate data structures course with 300+ students

WORK EXPERIENCE

Bregal Sagemount New York, NY
Technology Private Equity Associate, Flagship ($11B AUM) *Aug 2024 – Apr 2026*

- Sourced a **$200M+ transaction** for the investment in Pandion Optimization Alliance, a healthcare focused Group Purchasing Organization
- Developed LLM-powered automation workflows leveraging RAG, MCP integrations, and prompt engineering to automate outbound email drafting, investment sourcing, and meeting preparation for private equity investment professionals, increasing outbound volume by 14%+

Evercore New York, NY
Investment Banking Analyst, Technology M&A *Jul 2022 – Jun 2024*

- Built detailed operating and valuation models of private and public software, fintech, and IT services businesses for M&A transactions

Remee App New York, NY
Product Manager & Full-Stack Software Engineer Aug 2020 – Nov 2020

- Developed and shipped core features for a college social platform using React, JavaScript, and backend APIs, implementing user-facing workflows, integrating frontend and backend services; collaborated directly with founders to improve user engagement on the platform.

ADDITIONAL INFO

- **Languages:** Python, C/C++, Java, React / Javascript, HTML / CSS, R
- **Technical Frameworks:** PyTorch, Jax, vLLM, Hugging Face, Git, W&B, Slurm, Bash, SQL
- **Relevant Coursework:** Databases, Distributed Systems, Machine Learning, Robotics Research Seminar, Reinforcement Learning
- **Honors & Awards:** Egleston Scholar (Columbia Fellowship), Regeneron STS (Semifinalist), Siemens Competition (Semifinalist), AIME (3x)
- **Interests:** Running, Poker, Bouldering, Hiking, Chess, Catan, Tottenham Hotspur, Esports, Tennis, Piano, Cooking