

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

Masters of Science (Thesis-Based) in Computer Science (ML)

Advised by Professor Jiaqi Ma

Urbana, IL

Aug 2026 – Expected May 2022

Columbia University

Bachelor of Engineering in Computer Science (Intelligent Systems Track), Minor in Economics

New York, NY

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 & PROJECTS

TRAIS Lab at University of Illinois, Urbana-Champaign

Research Assistant under Jiaqi Ma

Remote

April 2026 – Present

Improbable AI Lab at Massachusetts Institute of Technology

Research Collaborator under Pulkit Agrawal

Remote

Apr 2025 – Present

- Leading data-centric machine learning research on novel methods to leverage data curation methods and synthetic data generators to efficiently train large language models (LLMs) to enable useful inductive biases for visual reasoning and general language modeling tasks; responsible for leading the ideation, execution, and design of all algorithms and experiments
- In [3], we show that pre-pre-training LLMs on synthetic data generated from neural cellular automata (NCAs) can improve the efficiency by up to 40% and perplexity by up to 6% on language modeling and improve model reasoning on GSM8K, HumanEval, and Bigbench tasks with 10x less compute and data than pre-pretraining on natural language; we also find that the optimal complexity for NCA pre-pre-training data varies depending on the downstream domain enabling data curation based on gzip compressibility

Multimodal Learning on Decision Transformers for Adversarial Robustness

Aug 2021 - Dec 2021

- Developed a multimodal Decision Transformer (DT) framework integrating proprioception, RGB, and depth data to learn vision-based robotic policies and improve adversarial robustness in offline reinforcement learning as part of a course project in Robotics Learning
- Implemented white-box (FGSM) and black-box (Gaussian noise) adversarial attacks on vision-based robotic policies to evaluate robustness
- Engineered and trained a convolutional variational autoencoder (CVAE) to extract latent embeddings from RGB and depth images to generate sequential inputs for the decision transformer model
- Conducted ablation studies to analyze the impact of different modality-based adversarial attacks on policies learned for the Franka Kitchen D4RL task; experiments demonstrated the robustness of multimodal DT models compared to single modal models

Burke Energy & Combustion Lab

NSF Undergraduate Research Intern

New York, NY

Jul 2019 – Aug 2019; May 2020 – Jul 2020

- Developed programs in Python to mine data from research journal graphs for uncertainty quantification models through a combination of image pre-processing, machine learning, template-based correlation matching, and clustering methods
- Assisted graduate students in developing, refactoring, and improving codebase efficiency for simulations and quantitative models

PUBLICATIONS & PREPRINTS

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

[2] Liu, A., Liu, B., Lee, D., Weissman, M., Posner, J., Cha, J., & Yoo, S. (2017). Machine learning aided prediction of family history of depression. Proceedings of the 2017 IEEE New York Scientific Data Summit (NYSDS). doi.org

[1] Lee, D., Cisek, D., Yoo, S. (2017). Sensor-network-based Wind Field Estimation using Deep Learning. Proceedings of the 2017 IEEE New York Scientific Data Summit (NYSDS). doi.org

TEACHING EXPERIENCE

Columbia University

Head Teaching Assistant

Teaching Assistant

New York, NY

Fall 2021

Fall 2020 - Fall 2021

- Led and organized homework and exam grading, and recitations for Columbia's largest undergraduate data structures course (300+ students), COMS 3134 - Data Structures, which covered basic data structures (stacks, queues, graphs), algorithms, and implementations in an object-oriented programming language (Java)

WORK EXPERIENCE

Bregal Sagemount

Technology Private Equity Associate, Flagship (\$11B AUM)

New York, NY

Aug 2024 – Apr 2026

- Prepare IC memos & operating models for buyouts & minority investments in software & services assets ranging from \$200M – \$1B TEV
- Sourced a majority buyout in Pandion Optimization Alliance, a leading Healthcare group purchasing organization
- Developed programs to leverage popular LLMs to automate outbound email drafting and sourcing tasks for investment professionals

Evercore*Investment Banking Analyst, Technology M&A*

New York, NY

*Jul 2022 – Jun 2024**Investment Banking Summer Analyst, Generalist**Jun 2021 – Aug 2021*

- Developed detailed operating and valuation models of private and public software, fintech, and IT services businesses for M&A transactions
- Advised Commscope (NASDAQ: VSN) on carve-outs of core business segments, including the sale of the Home business to Vantiva

Remee App

New York, NY

*Product Manager**Aug 2020 – Nov 2020*

- Worked directly with start-up founders and software engineers to develop a college-oriented social media platform and application
- Managed an agile team and oversaw development and testing sprints in preparation for investor demos and IOS app launch

ADDITIONAL INFO

- **Programming Languages & Frameworks:**
 - **Proficient:** Python (Jax / PyTorch), C / C++, Java, Huggingface, Wandb, Slurm
 - **Familiar:** React, Javascript, HTML / CSS, SQL, Bash / Shell
- **High School:** Commack High School (June 2018), Regeneron STS (Semifinalist), Siemens Competition (Semifinalist), AIME (3x)
- **Interests:** Running, Poker, Boulderling, Hiking, Chess, Catan, Tottenham Hotspur, Esports, Tennis, Piano, Cooking, Competition Math / Programming