

Jin Hwa Lee

3rd year PhD student @ Theory of Learning Lab, University College London

✉ jin.lee.22@ucl.ac.uk 🏠 [jinhl9.github.io](https://github.com/jinhl9) | 📄 [google scholar](#)

Sainsbury Wellcome Centre & Gatsby Computational Neuroscience Unit, 25 Howland St. London W1T 4JG

RESEARCH STATEMENT

My research focuses on understanding how the structure of data and the inductive biases of models shape learning. I believe that a fundamental scientific understanding of learning is essential for explaining the surprising capabilities of current artificial intelligence systems, including language production and reasoning, and to ultimately controlling them for reliable and efficient applications.

My work blends theory and controlled experiments based on tractable toy models, with empirical studies of models at scale. Through this approach, my current projects aim to understand how certain properties present in natural data interact with learning and generalization behavior of neural network models. In particular, I am interested in how various aspects of compositionality might emerge from this interplay.

EDUCATION

- **University College London** Oct 2022 - Present
PhD student
◦ Supervisor: Prof. Andrew Saxe London, UK
- **Technical University of Munich** Oct 2019 - Feb 2022
MSc Neuroengineering
◦ Supervisor: Prof. Mackenzie Mathis Munich, Germany
◦ Thesis: CEBRA: Multi-Modal Unsupervised Learning of Consistent Embeddings for Neural and Behavioral Activity
- **Korea Advanced Institute of Science and Technology(KAIST)** Mar 2015 - Sep 2019
BSc Physics Daejeon, South Korea
◦ Magna Cum Laude

RESEARCH

Lee, J. H., Lampinen, A., Singh, A.*, & Saxe, A.*, [Distinct Computations Emerge From Compositional Curricula in In-Context Learning](#), *Presented at ICLR 2025 SCSL Workshop, currently under review for full conference*.

- A demonstration of how curriculum-like data structures, richly present in natural language corpora, can influence models' in-context solution strategies on compositional tasks.

Lad, V., **Lee, J. H.**, Gurnee, W., Tegmark, M. [The Remarkable Robustness of LLMs: Stages of Inference?](#), *Under review*

- A framework for interpreting depth-dependent computations in LLMs.

Lee, J. H.*, Jiralerspong, T*, Yu, L., Bengio, Y., & Cheng, E., [Geometric Signatures of Compositionality Across a Language Model's Lifetime](#), *Accepted at ACL 2025 Main Conference*.

- Analyzing the geometric properties of hidden representations in LLMs throughout pretraining, and how the compositional structure of language is reflected in and correlated with the emergence of linguistic capabilities.

Dorrell, W.*, Hsu, K.*, Hollingsworth, L., **Lee, J. H.**, Wu, Jiajun., Finn, Chelsea., Latham, PE., Behrens, TEJ., & Whittington, JCR., [Range, not Independence, Drives Modularity in Biological Inspired Representation](#), *ICLR 2025*.

- Deriving necessary and sufficient conditions on sample data statistics for achieving modular representations under biological neural constraints.

Lee, J. H., Mannelli, S. S., & Saxe, A., [Why Do Animals Need Shaping? A Theory of Task Composition and Curriculum Learning](#), *ICML 2024*.

- Analytical study of deterministic policy learning dynamics of compositional RL in high-dimensional teacher-student setup.

Schneider, S.*, **Lee, J. H.***, & Mathis, M. W., [Learnable latent embeddings for joint behavioral and neural analysis](#), *Nature (2023)*.

- Contrastive learning and identifiability in ICA inspired multimodal ML method for mapping high dimensional neural and behavioral data.

Servadei, L., **Lee, J. H.**, Medina, J. A. A., Werner, M., Hochreiter, S., Ecker, W., & Wille, R., [Deep reinforcement learning for optimization at early design stages](#). *IEEE Design & Test (2022)*.

- Solving combinatorial optimization problem using pointer network model and reinforcement learning

INVITED TALKS

- **COSYNE 2025 Workshop: Compositional Learning**
Analytical Approach to Study Compositional Learning Apr 2025
Montreal, Canada
- **Invited talk: Learning Dynamics of Linguistic Compositionality**
Computational Linguistics Group, Universitat Pompeu Fabra, hosted by Marco Baroni & Emily Cheng Feb 2025
Barcelona, Spain
- **3rd Conference on Lifelong Learning Agents (CoLLAs)**
Tutorial: Theoretical Advances in Continual Learning, Itay Evron, Jin Hwa Lee Jul 2024
Pisa, Italy
- **COSYNE 2024 Workshop: Sharpening Our Sight**
CEBRA Tutorial Mar 2024
Cascais, Portugal
- **Invited talk: Tim Behrens group @ UCL, Oxford**
Analytical Model of Compositional Learning May 2023
London, UK

AWARDS AND SCHOLARSHIPS

- **Pivotal Research Fellowship for AI Safety** 2025
Pivotal, \$9500
- **Brain, Minds and Machines 2024 Summer School Travel Grant & Scholarship** 2024
Center for Brains, Minds and Machines, \$3000
- **COSYNE 2024 Travel Grant** 2024
COSYNE, \$1000
- **IEEE Brain BCI Hackathon** 2020
IEEE, 1st Prize
- **DAAD Scholarship** 2020
DAAD, \$ 13,000
- **National Science and Engineering Undergraduate Scholarship** 2017
KOSAF, \$ 11,000

TEACHING EXPERIENCE

- **Systems Neuroscience & Theoretical Neuroscience** Fall 2023
Sainsbury Wellcome Centre&Gatsby Computational Neuroscience Unit, Teaching Assistant London, UK
- **Machine Learning: Methods and Tools** Summer 2020
Technical University of Munich, Teaching Assistant Munich, Germany

OUTREACH & PROFESSIONAL DEVELOPMENT

- **Brains, Minds and Machines Summer School** Summer 2024
MIT CBMM, Participant Woods Hole, US
- **Women in Machine Learning Mentoring** 2023-2024
Mentor Remote
- **Analytical Connectionism** Summer 2023
Participant London, UK
- **Connect Foundation** 2016-2019
Education Volunteer Seoul, South Korea