

EDUCATION

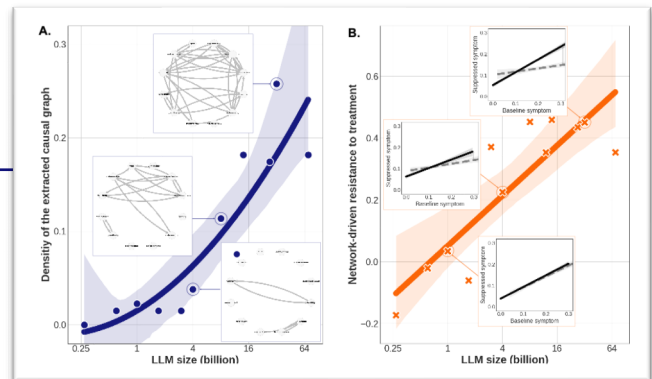
09/23-06/27 (expected)	Ph.D. in Artificial Intelligence, KAIST Thesis: “Modeling and Controlling AI Behavior and a Case of Machine Psychopathology” Advisor: Kijung Shin, Ph.D., Associate Professor of AI	
09/21-05/23	M.S. in Artificial Intelligence, KAIST Thesis: “Towards Deep Attention in Graph Neural Networks: Problems and Remedies” Advisor: Kijung Shin, Ph.D., Associate Professor of AI	
10/13-12/16	B.S. in Psychology, UCSD Thesis: “Mantram Repetition Program Reduces Depression Severity among Veterans with Major Depressive Disorder” Advisor: Sidney Zisook, M.D., Distinguished Professor of Psychiatry	GPA: 3.84/4.00

HONOR & AWARD

2026	Winner of Global AI Research Leap Challenge, hosted by National AI Research Lab of Korea
2025	The outstanding reviewers of KDD 2025 (top 10%)
2024	The best reviewers of KDD 2024 (top 1.2%; 30/2558)
2016	<i>magna cum laude</i> , UCSD
2014-2016	Psychology Honors, UCSD
2014-2016	Faculty Mentor Program, UCSD

FEATURED WORK

In my previous work [[link](#)], I collaborated with [a leading psychologist](#) to provide the first evidence that computations of psychopathology (i.e., mental disorders) may have emerged in LLMs. We computationally adapted [a leading theory](#) about the nature of psychopathology, and tested the theory in LLM using mechanistic interpretability methods and behavioral simulations across 12 different LLMs.



Computational structure of psychopathology (*left*) and its resulting dysfunction (*right*) scale with LLM size.

TALK & TUTORIAL

- [T2] Emergence of psychopathological computations in large language models. Hosted by Prof. Denny Borsboom, *Faculty of Social and Behavioural Sciences, UvA*. (Feb., 2026).
- [T1] Hypergraph Neural Networks: An In-depth and Step-By-Step Guide. In *KDD2024, ICDM2024, AAAI2025, CIKM2025*.

PUBLICATION & PREPRINT

* denotes equal contribution

- [P2] Lee, S. Y.*, Lee, J.*, Chun, J., Hwang, H., Bu, F., Ben-Zion, Z., Kim, T., Borsboom, D., Yoo, J., & Shin, K. (2026). Automating and Scaling Behavioral Scientific Research on AI Agents. In *arXiv* preprint arXiv:2608.10030. [[link](#)]
- [C13] Kang, S., Lee, S. Y., Kim, J., Shin, K., & Jang, B. (2026). Improving Rare Medication Recommendation with Counterfactual Data Augmentation and Large Language Models. In *RecSys*. [[link](#)]
- [C12] Lee, G.*, Bu, F.*, Lee, S. Y.*, Kim, S., & Shin, K. (2026). Toward a Science of AI Agent Societies. In *KDD*. [[link](#)]
- [C11] Kim, S., Lee, S. Y., Kim, K., Hwang, H., Yoo, J., & Shin, K. (2026). Feature-Centric Unsupervised Node Representation Learning Without Homophily Assumption. In *AAAI*. [[link](#)]
- [C10] Kim, S., Lee, S. Y., Yoo, J., & Shin, K. (2025). 'Hello, World!': Making GNNs Talk with LLMs. In *EMNLP Findings*. [[link](#)]

- [P1] Lee, S. Y., Hwang, H., Kim, T., Kim, Y., Park, K., Yoo, J., Borsboom, D., & Shin, K. (2025). Emergence of psychopathological computations in large language models. In *arXiv* preprint arXiv:2504.08016. [\[link\]](#)
[†] *featured as part of presidential address in IMPS 2025 by Professor Denny Borsboom*
- [C9] Jo, H., Hwang, H., Bu, F., Lee, S. Y., Park, C., & Shin, K. (2025). On Measuring Unnoticeability of Graph Adversarial Attacks: Observations, New Measure, and Applications. In *KDD*. [\[link\]](#)
- [W1] Yu, T., Lee, S. Y., Hwang, H., & Shin, K. (2024). Prediction Is NOT Classification: On Formulation and Evaluation of Hyperedge Prediction. In *ICDM-W*. [\[link\]](#)
- [C8] Kim, S., Lee, S. Y., Bu, F., Kang, S., Kim, K., Yoo, J., & Shin, K. (2024) Rethinking Reconstruction-based Graph-Level Anomaly Detection: Limitations and a Simple Remedy. In *NeurIPS*. [\[link\]](#)
- [J4] Kim, M., Lee, J., Lee, S. Y., Ha, M., Park, I., Jang, J., ... & Kwon, J. S. (2024). Development of an eye-tracking system based on a deep learning model to assess executive function in patients with mental illnesses. *Scientific Reports*, 14(1), 18186. [\[link\]](#)
- [C7] Kim, S.*, Lee, S. Y.*, Gao, Y., Antelmi, A., Polato, M., & Shin, K. (2024). A Survey on Hypergraph Neural Networks: An In-Depth and Step-By-Step Guide. In *KDD*. [\[link\]](#)
- [C6] Bu, F., Jo, H., Lee, S. Y., Ahn, S., & Shin, K. (2024). Tackling Prevalent Conditions in Unsupervised Combinatorial Optimization: Cardinality, Minimum, Covering, and More. In *ICML*. [\[link\]](#)
- [C5] Lee, S. Y., Kim, S., Bu, F., Yoo, J., Tang, J., & Shin, K. (2024). Feature Distribution on Graph Topology Mediates the Effect of Graph Convolution: Homophily Perspective. In *ICML*. [\[link\]](#)
[†] *featured in the Research Webzine, KAIST Breakthroughs Vol. 24* [\[link\]](#)
- [C4] Lee, G., Lee, S. Y., & Shin, K. (2024). Villain: Self-Supervised Learning on Homogeneous Hypergraphs without Features via Virtual Label Propagation. In *WWW*. [\[link\]](#)
- [C3] Kim, S., Kang, S., Bu, F., Lee, S. Y., Yoo, J., & Shin, K. (2024). HypeBoy: Generative Self-Supervised Representation Learning on Hypergraphs. In *ICLR*. [\[link\]](#)
- [C2] Lee, S. Y*, Kim, J*, Park, K., Ryu, D. K., Shim, S., & Shin, K. (2023). You're Not Alone in Battle: Combat Threat Analysis Using Attention Networks and a New Open Benchmark. In *CIKM*. [\[link\]](#)
- [C1] Lee, S. Y., Bu, F., Yoo, J., & Shin, K. (2023). Towards Deep Attention in Graph Neural Networks: Problems and Remedies. In *ICML*. [\[link\]](#)
- [J3] Pospos, S., Tal, I., Iglewicz, A., Newton, I. G., Tai-Seale, M., Downs, N., Jong, P., Lee, D., Davidson, J. E., Lee, S. Y., Rubanovich, C. K., Ho, E. V., Sanchez, C., & Zisook, S. (2019). Gender differences among medical students, house staff, and faculty physicians at high risk for suicide: A HEAR report. *Depression and anxiety*, 36(10), 902-920. [\[link\]](#)
- [J2] McClure, J., Macera, C., Ji, M., Nievergelt, C., Lee, S. Y., Kayman, J., & Zisook, S. (2018). Agreement between self and psychiatrist reporting of suicidal ideation at a Veterans Administration psychiatric emergency clinic. *Depression and anxiety*, 35(11), 1114-1121. [\[link\]](#)
- [J1] Worley, M. J., Isgro, M., Heffner, J. L., Lee, S. Y., Daniel, B., & Anthenelli, R. M. (2018). Predictors of reduced smoking quantity among recovering alcohol dependent men in a smoking cessation trial. *Addictive behaviors*, 84, 263-270. [\[link\]](#)

WORK EXPERIENCE

09/2021 – Present	Graduate Researcher at Data Mining Lab Advisor: Kijung Shin, Ph.D. (Associate Professor of AI, KAIST)
09/2020 – 06/2021	Clinical Research Assistant at Clinical Cognitive Neuroscience Center Advisor: Jun Soo Kwon, M.D., Ph.D. (Professor of Psychiatry, Seoul National University)
10/2018 – 05/2020	Sergeant Operation Agent at the Republic of Korea Army
08/2018 – 09/2018	Clinical Research Assistant at Severance Laboratory of Early Life Stress Advisor: Jeong Ho Seok, M.D., Ph.D. (Professor of Psychiatry, Yonsei University)
01/2016 – 02/2018	Clinical Research Assistant at San Diego VA Hospital Research Department Advisor: Sidney Zisook, M.D. (Distinguished Professor of Psychiatry, UCSD)
01/2016 – 06/2016	Clinical Research Assistant at San Diego VA Hospital Research Department

06/2015 – 08/2015 Advisor: Ariel Lang, Ph.D., M.P.H. (Professor of Psychiatry, UCSD)
Clinical Research Assistant at UCSD Psychiatry Department
 Advisor: Hagop Akiskal, M.D. (Emeritus Professor of Psychiatry, UCSD)

03/2015 – 03/2016 **Clinical Research Assistant** at Pacific Treatment and Research Center
 Advisor: Robert Anthenelli, M.D. (Professor of Psychiatry, UCSD)

SKILLS

Category	Experience	Representative Works
Domain expertise	AI/ML Psychology/Cognitive science/Psychiatry Network science Philosophy of mind	P1, C1, C3, C5 P1, P2, J1-J4 C1, C4, C5 P1
Research	Engineering research Scientific research Clinical research Theoretical research Behavioral research Large scale research (Randomized) Controlled experiment Simulation experiment	C1, C3 P1, P2, C5 J1-J4 P1, C1, C5 P1, P2, J4 P1, P2, C5, C9 P1, P2, C5, J1 P1, P2, C5
Deep learning methods	Transformers/Large language models Graph neural networks Autoencoders	P1, P2, C10 C1, C3, C5, C7 P1, C8
Method design	Multi-agent system design Algorithm design Measure design Self-supervised loss design	P2 P1, C1, C5 P1, C5, C9 P1, C3, C4
Analysis	Mechanistic interpretability Causal inference Statistical analysis Time series analysis Network analysis (Interpretable) Computational modeling Synthetic data generation	P1 P1, P2, C5 P1, P2, C5 P1 P1, C5 P1, C5 P1, C5
Application	Research automation Adversarial attack Anomaly detection Healthcare Military	P2 P1, C9 C8 J1-J4 C2
Teamwork	Communication Cross-institutional and -disciplinary collaboration	T1, T2, Best reviewer awards P1, P2, C2, C7, J4
Natural language	English (bilingual proficiency) Korean (native) Chinese (limited working proficiency)	