Gouki Minegishi

Tokyo, Japan
E-mail: minegishi@weblab.t.u-tokyo.ac.jp

EDUCATION

Bachelor of Engineering, University of Tokyo（2019~2023）
Master of Engineering, University of Tokyo（2023~2025）
PhD of Engineering, University of Tokyo（2025~）

RESEARCH/WORK EXPERIENCE

- Research Intern, Preferred Networks, Inc （August 2024 – October 2024）
  - Research on in-context learning in foundation models, with a particular focus on knowledge conflicts.
- Chief AI Engineer, Matsuo institute, Inc （November 2021- present）
  - Developed automated driving software, including 3D object detection with deep learning and integration of the World Model into Automated Driving.
- AI Engineer , GHELIA, Inc （April 2021 – March 2022）
  - Specialized in risk prediction algorithms and models.

RESEARCH INTERESTS

- Mechanical Interpretability
- In Context Learning

SELECTED PUBLICATIONS

- **Minegishi, G.**, Furuta, H., Taniguchi, S., Iwasawa, Y., & Matsuo, Y. .
  “Beyond Induction Heads: In-Context Meta Learning Induces Multi-Phase Circuit Emergence.”
  International Conference on Machine Learning (ICML 2025).
- **Minegishi, G.**, Furuta, H., Iwasawa, Y., & Matsuo, Y..
  “Rethinking Evaluation of Sparse Autoencoders through the Representation of Polysemous Words.” International Conference on Learning Representations *(ICLR 2025).*
- Taniguchi, S., Harada, K., **Minegishi, G.**, et al.
  “ADOPT: Modified Adam Can Converge with Any β2 with the Optimal Rate.”
  Neural Information Processing Systems *(NeurIPS 2024).*
- Furuta, H., **Minegishi, G.**, Iwasawa, Y., & Matsuo, Y..
  “Towards Empirical Interpretation of Internal Circuits and Properties in Grokked Transformers on Modular Polynomials.”
  Transactions on Machine Learning Research *(TMLR)*.
- **Minegishi, G**., Iwasawa, Y., & Matsuo, Y.
  “Bridging Lottery Ticket and Grokking: Understanding Grokking from Inner Structure of Networks.”
  Transactions on Machine Learning Research *(TMLR)*.

ACADEMIC HONORS

- Selected for the "BOOST NAIS" Scholarship Program for fostering advanced AI talents leading the next-generation intelligent society
- Young Researcher Encouragement Award, NLP 2025
- Outstanding Presentation Award, JSAI Annual Conference 2024
- Oral presentation accepted at the ICLR 2024 Workshop on Bridging the Gap Between Practice and Theory in Deep Learning

ACADEMIC ACTIVITY

- Co-organizer for **Mechanistic Interpretability Organized Session** at JSAI2025.