HAOLIN YANG

✉ haolinyang2001@uchicago.edu ☎+1 7732567911 ◇ Personal Website

EDUCATION

University of Chicago **Sept. 2024 - Now**
Master of Science in Statistics (GPA: 3.96/4.0)
Courses: Introduction to Computer Vision, Stochastic Calculus, Geometric Methods in Computer Science

Tsinghua University **Sept. 2020 - June 2024**
Bachelor of Arts in English (GPA: 3.94/4.0)
Minor in Economics and Finance (GPA: 4.0/4.0)
Courses: Nonparametric Statistics, Linear Regression Analysis, Real Analysis, Stochastic Processes

RESEARCH INTERESTS

In-Context Learning, Mechanistic Interpretability, Large Language Models

PUBLICATIONS AND MANUSCRIPTS

1. **Yang, H.**, Cho, H., Zhong, Y., & Inoue, N. (2025). Unifying Attention Heads and Task Vectors via Hidden State Geometry in In-Context Learning. Advances in Neural Information Processing Systems 39 **(NIPS)**, 2025. [Link]

2. **Yang, H.**, Cho, H., & Inoue, N. (2025). Localizing Task Recognition and Task Learning in In-Context Learning via Attention Head Analysis. Submitted to **ICLR 2026**. [Link]

3. **Yang, H.**, Cho, H., Ding, K., & Inoue, N. (2025). Task Vectors, Learned Not Extracted: Performance Gains and Mechanistic Insight. Submitted to **ICLR 2026**. [Link]

4. Cho, H., **Yang, H.**, Minegishi, G., & Inoue, N. (2025). Mechanism of Task-Oriented Information Removal in In-Context Learning. Submitted to **ICLR 2026**. [Link]

5. Cho, H., **Yang, H.**, Kurkoski, B. M., & Inoue, N. (2025). Binary Autoencoder for Mechanistic Interpretability of Large Language Models. Submitted to **ICLR 2026**. [Link]

6. Zhang, Y., Yang, F., **Yang, H.**, & Han, S. (2024). Does Checking-In Help? Understanding L2 Learners' Autonomous Check-In Behavior in an English-Language MOOC through Learning Analytics. **ReCALL**, 36(3), 343-358. [Link]

RESEARCH EXPERIENCES

Localizing Task Recognition and Task Learning in In-Context Learning **June 2025 - Sep. 2025**
Advisor: Naoya Inoue
- Proposed the Task Subspace Logit Attribution method to identify two types of attention heads in Transformer LLMs responsible for task learning and task recognition in in-context learning.
- Analyzed how the heads independently and effectively govern task learning and task recognition through ablation, steering, and input perturbation experiments.
- Revealed how the two types of heads collectively achieves the ICL functionality by investigating their contributions to the hidden states updates from a geometric perspective.

Task Vectors in In-Context Learning and their Mechanism **May 2025 - Sep. 2025**
Advisor: Kaize Ding, Naoya Inoue
- Proposed a method to directly train task vectorS in in-context learning which outperforms previous methods based on distilling LLM hidden states in terms of flexibility, scalability, and adaptability.
- Revealed the interaction between task vectors and OV circuits of attention heads as the predominant factor in the low-level mechanism by which task vectors influence model outputs.

- Discovered a strongly linear pattern in the high-level mechanism by which task vectors influence model outputs through analyzing LLM layer updates using numerical and spectral methods.

Geometric Analysis of the LLM Hidden States under ICL setting **July 2024 - May 2025**
Advisor: Yiqiao Zhong, Naoya Inoue
- Provided a two-stage geometric characterization of the layer-wise evolution of LLM hidden states of ICL prompts using quantitative measures, with the early layers promoting separability and late layers advancing the alignment of hidden states with label words' unembeddings.
- Demonstrated the correlation between the two-stage pattern and the semantic dynamics of hidden states using spectral analysis, and proposed a method to denoise hidden states for better prediction accuracy.
- Analyzed how previous token heads and induction heads as two special types of attention heads contribute to the two stages respectively, and revealed how task vectors as a explanation for LLM's ICL functionality fits into the two-stage framework.

Automatic Post-Editing of English-Chinese Machine-Translated Academic Texts **Jan. 2024 - June 2024**
- Designed and fine-tuned an Bert-based automatic post-editing model using ParaMed, a English-Chinese parallel corpus containing academic texts in the field of biomedicine.
- Performed comparative statistical analyses using the BLEU and TER metrics to validate the quality improvements brought by the post-editing model over the raw machine-translated texts.
- Qualitatively examined the post-editing strategy of the model by investigating how it corrects the machine translation errors and its failure scenarios.

Check-in Behaviors and Academic Performance in MOOC **March 2021 – May 2024**
- Collected and cleaned the academic performance data of 11296 students of an English MOOC on the XueTangX platform including their check-in instances and scores in assignments and tests.
- Conducted chi-square test of independence and logistic regression to reveal to what extent students' check-in behaviors affect their course completion rates.
- Conducted Mann-Whitney U tests between the students who checked in regularly and those who did not check in to determine the differences in their learning behaviors.

HONOR AND AWARDS

Academic Excellence Scholarship, Tsinghua University (2022-2023)
Outstanding Undergraduate, Tsinghua University (2024)
Outstanding Undergraduate Thesis, School of Humanities, Tsinghua University (2024)

SKILLS

Computer Programming

- Coding Languages: Proficient in Python, R, Stata.
- Deep Learning Frameworks: PyTorch, Transformers.

Languages

- English: TOEFL 117 (Reading 30; Listening 30; Writing 30; Speaking 27) GRE 339+5.0 (Verbal 169; Quantitative 170; AW 5.0)
- Mandarin: Native
- German: C1 (Proficient)