

ZHONGYU OUYANG

Ph.D. student in Computer Science

Email: zhongyu.ouyang.gr@dartmouth.edu

Phone: (412) 320-0590

Homepage: <https://zyouyang.github.io/> [GitHub](#) [LinkedIn](#) [Google Scholar](#)

Research Interests

Large Language Models (LLMs)-based Recommender Systems; LLM Preference Optimization with Cognitive Science; Ranking; Graph Learning

Skills

Python (PyTorch, Hugging Face, Deep Graph Library, etc.), C/C++/C#, Unix Shell, SQL/SCOPE, Git, Latex, JAX

Education

Dartmouth College

Hanover, NH, US

Ph.D. Student in Computer Science Sep. 2024 – June. 2026 (Expected)

University of Notre Dame

South Bend, IN, US

Ph.D. Student in Computer Science Aug. 2021 – May. 2024

Transferred to Dartmouth College; GPA: 4.0/4.0

Georgia Institute of Technology

Atlanta, GA, US

M.S. in Computer Science Jan. 2020 – Aug. 2022

Minor in Machine Learning; GPA: 3.9/4.0

University of Pittsburgh

Pittsburgh, PA, US

M.S. in Industrial Engineering Aug. 2018 – Apr. 2020

Tianjin University (#260 in U.S. News Global)

Tianjin, CHN

B.E. in Material Science Sep. 2014 – July 2018

Experience

Bing Places, Microsoft, Redmond, WA

PhD Data Scientist Intern

May 2024 – Aug. 2024

Optimized auto-suggestion system for complex address lines of local chain businesses in international regions, improving user experience & accuracy:

– Led two strategic **LLM-powered Search Engine initiatives**:

- Developed an LLM-based solution leveraging advanced prompt engineering to automatically extract critical address components using the model's inherent world knowledge.

- Designed a token-based classification approach utilizing the AddressParser model to systematically identify and categorize address elements with high precision.

– Integrated AddressParser into the existing production pipeline through cross-platform implementation using C#, SCOPE, and Cosmos, enabling explicit extraction of address components from unstructured address lines.

Qeexo, Pittsburgh, PA

Machine Learning Engineer

June 2020 – Sep. 2020

- Integrated ML models with embedded pipelines to support the back-end of the Qeexo AutoML application designed for automatic machine learning support for embedded hardware and systems [\[demo\]](#).
- Lead the project in sound recognition with Qeexo AutoML and covered by company’s blog [\[blog\]](#).
- Research on the applicability of the state-of-the-art hyperparameter tuning methods such as successive halving to improve model performance, computation time, and space consumption. For linear SVM, the space consumption is decreased by at least 50%, and the computation time is decreased by 40%.

Research Projects

Aligning LLMs with Human-based Preferences from Cognitive Science’s Perspective Winter 2025

- Modeling preference hierarchies for human-aligned sequential recommendations with LLMs (Llama 3-8B).
- Adapted recommendation agent ranks items closely adheres to user preferences while avoiding undesirable ones, sustaining user engagement.
- Preprint [\[link\]](#) (submitted to ACL 2025):

Prompt Learning over the Mobile Application Network in the Wild with Transformers Spring 2023

- Abstract app features and promotion behaviors as a heterogeneous graph
- Pioneered the incorporation of both *embedding-based* and *metapath-based* prompt tuning techniques with a pre-trained BERT model, providing pre-attentive information guidance within the app promotion network.
- The work (paper, code, and dataset) is available at [\[link\]](#) and the related paper is published under ECML PKDD’24.

Dual Representation Alignment and Uniformity in Recommender Systems Fall 2022

- Calibrate the trade-off between representation alignment and uniformity through dually supervised and self-supervised tasks, enabling a robust and fast training process.
- Conduct extensive experiments on three datasets to demonstrate improved model performance and convergence speed, along with further training dynamic analysis.
- The paper first appeared at the ICML 2023 workshop and is published under ICWSM’24.

Papers

From Thoughts to Tastes: Modeling Preference Hierarchies for Human-Aligned Sequential Recommendations [\[PDF\]](#) [\[Code\]](#)

{Q. Wen*, Z. Ouyang*}, C. Zhang, S. Vosoughi, Y. Ye

**Equal Contribution and Random order*

Scaled Supervision is an Implicit Lipschitz Regularizer [\[PDF\]](#) [\[Code\]](#)

Z. Ouyang, C. Zhang, Y. Jia, S. Vosoughi

Under Review

Leveraging Graphs to Click-Through Rate Prediction through Test-Time Augmentation [PDF]

Z. Ouyang, M. Ju, Y. Ye

Under Review

On the Influence of Connectivity to CTR Prediction [PDF]

Z. Ouyang, C. Zhang, Y. Jia, S. Vosoughi

Under Review

Pretrained Image-Text Models are Secretly Video Captioners [PDF] [Code]

C. Zhang, Y. Jian, **Z. Ouyang**, S. Vosoughi

NAACL'25 Main: Annual Conference of the North American Chapter of the Association for Computational Linguistics

Temporal Working Memory: Query-Guided Temporal Segment Refinement for Enhanced Multimodal Understanding [PDF] [Code]

X. Diao*, C. Zhang*, W. Wu, **Z. Ouyang**, P. Qing, M. Cheng, S. Vosoughi, J. Gui

NAACL'25 Findings: Annual Conference of the North American Chapter of the Association for Computational Linguistics

Symbolic Prompt Tuning Completes the App Promotion Graph [PDF] [CODE]

Z. Ouyang, S. Hou, S. Ma, C. Chen, C. Zhang, Y. Ye et al.

ECML PKDD'24: European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases

How to Improve Representation Alignment and Uniformity in Graph-based Collaborative Filtering? [PDF] [CODE]

Z. Ouyang, S. Hou, C. Zhang, C. Zhang, Y. Ye

ICWSM'24: Proceedings of the International AAAI Conference on Web and Social Media

Working Memory Identifies Reasoning Limits in Language Models [PDF]

C. Zhang, Y. Jian, **Z. Ouyang**, S. Vosoughi

EMNLP'24: The Conference on Empirical Methods in Natural Language Processing

Disentangled Heterogeneous Dynamic Graph Learning for Opioid Overdose Prediction [PDF]

Q. Wen, **Z. Ouyang**, J. Zhang, Y. Qian, C. Zhang, Y. Ye

KDD'22: SIGKDD Conference on Knowledge Discovery and Data Mining

Service

Teaching

COSC 294 - Reading Course for M.S. in Digital Arts

Conference Reviewer

ICLR 2024, NeurIPS 2023, ICML 2023, KDD 2023, AAAI 2024, ICDM 2023.

Journal Reviewer

IEEE Transactions on Dependable and Secure Computing, IEEE Transactions on Big Data.

Workshop Reviewer

NeurIPS 2023 Workshop: Socially Responsible Language Modeling Research (SoLaR2023), 2nd KDD Workshop on Uncertainty Reasoning and Quantification in Decision Making (UDMKDD2023), 2nd ACM SIGKDD Workshop on Ethical Artificial Intelligence: Methods and Applications (EAIKDD2023), 2nd Workshop on Formal Verification of Machine Learning (WVVML2023), The First Workshop on DL-Hardware Co-Design for AI Acceleration (DCAA2023)