Xunjian Yin

+1-919-423-3222 | xunjian.yin@duke.edu | xunjianyin.github.io

linkedin | github | scholar | twitter

Duke University, Durham - 27705, US.

Research Interests

#Agent #World Model #Continuous Learning #RL

Education

- **Duke University** *08.2025 - 06.2029 (expected)*
  Ph.D. in Computer Science, supervised by Prof. Shuyan Zhou
- **Peking University** *09.2022 - 06.2025*
  M.S. in Computer Science, supervised by Prof. Xiaojun Wan GPA: 3.81/4.00
- **Peking University** *09.2018 - 06.2022*
  B.S. in Computer Science GPA: 3.64/4.00

Selected Publications

Preprints:

[1] **LEDOM: An Open and Fundamental Reverse Language Model** *Under review* [link]
Xunjian Yin, Sitao Cheng, Yuxi Xie, Xinyu Hu, Li Lin, Xinyi Wang, Liangming Pan, William Wang, Xiaojun Wan.

[2] **The Geometry of Reasoning: Flowing Logics in Representation Space** *Under review* [link]
Yufa Zhou*, Yixiao Wang*, **Xunjian Yin***, Shuyan Zhou, Anru R. Zhang.

[3] **ContraSolver: Self-Alignment of LLMs by Resolving Internal Preference Contradictions** *Under review* [link]
Xu Zhang*, **Xunjian Yin***, Xiaojun Wan.

Conference paper:

[1] **DAMON: A Dialogue-Aware MCTS Framework for Jailbreaking Large Language Models**
EMNLP 2025 [link]
Xu Zhang, **Xunjian Yin**, Dinghao Jing, Huixuan Zhang, Xinyu Hu, Xiaojun Wan.

[2] **Gödel Agent: A Self-Referential Agents Framework for Recursively Self-Improvement**
ACL 2025 [link]
Xunjian Yin, Xinyi Wang, Liangming Pan, Xiaojun Wan, William Wang.

[3] **ChemAgent: Self-updating Memories in LLMs Improves Chemical Reasoning**
ICLR 2025 [link]
Xiangru Tang, Tianyu Hu, Muyang Ye, Yanjun Shao, **Xunjian Yin**, ..., Arman Cohan, Mark Gerstein

[4] **Benchmarking Knowledge Boundary for LLMs: A Different Perspective on Model Evaluation**
ACL 2024 [link]
Xunjian Yin*, Xu Zhang*, Jie Ruan, Xiaojun Wan.

[5] **History Matters: Temporal Knowledge Editing in Large Language Models**
AAAI 2024 [link]
Xunjian Yin, Jin Jiang, Liming Yang, Xiaojun Wan.

[6] **Error-Robust Retrieval for Chinese Spelling Check**
COLING 2024 [link]
Xunjian Yin, Xinyu Hu, Jin Jiang, Xiaojun Wan.

[7] **ALCUNA: Large Language Models Meet New Knowledge**
EMNLP 2023 [link]
Xunjian Yin*, Baizhou Huang*, Xiaojun Wan.

[8] **How Do Seq2Seq Models Perform on End-to-End Data-to-Text Generation?**
ACL 2022 [link]
Xunjian Yin, Xiaojun Wan.

[9] **Themis: A Reference-free NLG Evaluation Model with Flexibility and Interpretability**
EMNLP 2024 [link]
Xinyu Hu, Li Lin, Mingqi Gao, **Xunjian Yin**, Xiaojun Wan.

[10] **DSGram: Dynamic Weighting Sub-Metrics for Grammatical Error Correction in the Era of LLMs**
AAAI 2025 [link]
Jinxiang Xie, Yilin Li, **Xunjian Yin** (as Mentor), Xiaojun Wan.

Research Experience

- **University of California, Santa Barbara (NLP Group)** *06.2024 - 10.2024*
  ***Advisor**: Prof. William Wang* ***Role**: Visiting Research Scholar* California, USA
  - **Project 1**: Reverse Language Model (RLM) Pre-training, Evaluation, Analysis and Applications
    - We trained RLMs (2B-7B) from scratch, using a 500 billion tokens and the last-token prediction approach.
  - **Project 2**: Gödel Agent: A Self-Referential Agents Framework for Recursively Self-Improvement
    - We developed a self-referential agent, which is capable of reading and modifying its own logic and code.
- **Microsoft Research Asia (NLC Group)** *02.2022 - 08.2022*
  ***Advisor**: Dr. Shuming Ma and Dr. Kai Chen* ***Role**: Research Intern* Beijing, China
  - **Project:** Pre-training with Curriculum Learning (CL)
    - We investigated the application of CL during pretraining, which accelerates the convergence of the pretraining.
- **Wangxuan Institute of Computer Technology, Peking University** *10.2020 - 06.2022*
  ***Advisor**: Prof. Xiaojun Wan* ***Role**: Research Assistant* Beijing, China
  - **Project 1**: Analysis of Seq2Seq Models on End-to-End Data-to-Text Generation
    - We find that the most advanced models do not always yield the best performance and larger models are better.
  - **Project 2:** Enhancing Language Models with kNN for Grammar Error Correction (GEC)
    - A kNN algorithm with robust information derived from character phonetics and shapes to augment LMs.
- **Institute of Computational Linguistics, Peking University** *04.2020 - 11.2021*
  ***Advisor**: Prof. Yunfang Wu* ***Role**: Research Assistant* Beijing, China
  - **Project:** Multi-Task Learning for Grammar Error Correction
    - We introduced a dependency tree recovery task to enhance the performance of the BART model in grammar.
- **Institute of Computational Linguistics, Peking University** *07.2019 - 12.2019*
  ***Advisor**: Prof. Sujian Li* ***Role**: Research Assistant* Beijing, China
  - **Project:** Building Benchmark for Mathematical Olympiad Problems
    - We create a benchmark using middle-school-level mathematical Olympiad problems as the testbed.

Honors and Awards

- **Outstanding Graduate of Beijing,** Beijing Municipality *06.2025*
- **Outstanding Graduate of Peking University,** Peking University *06.2025*
- **Outstanding Master's Thesis Award of WICT,** Peking University *06.2025*
- **Merit Student,** Peking University *09.2024*
- **Guotai Junan Scholarship,** Peking University *09.2024*
- **Wang Xuan Scholarship,** Peking University *09.2022*
- **Award for Research Excellence,** Peking University *09.2021, 09.2022, 09.2023*
- **Award for Academic Excellence,** Peking University *09.2019, 09.2020*
- **Outstanding Student of Shandong Province,** Shandong Province *04.2018*

Service

- **Teaching Assistant:** *2021 - Present*
  Peking University
  - Introduction to Computing (C++, 2021 fall)
  - Data Structures and Algorithms (2022 spring)
  - Introduction to Computing (Python, 2023 fall)
  - Web Data Mining (2023 fall)
- **Reviewer:** *2022 - Present*
  - ACL-ARR (2023 - Now), ICLR (2024 - 2026), NeurIPS (2024 - 2025), ICML (2025), AAAI (2024 - 2025), COLM (2025), COLING (2024), JCST (2025)
- **Volunteer:** *2023 - Present*
  - AAAI'24, ACL'24, NLPCC'23 Shared Task 8 track chair

Skills

- Pytorch, Transformers, Distributed Training, Accelerated Inference
- Python, C++, HTML, JavaScript