Xunjian Yin

+86-131-4630-1177 | xjyin@pku.edu.cn | xunjianyin.github.io

linkedin | github | scholar | twitter

Peking University, Beijing - 100871, China

Research Interests

#LLM Evaluation #Agent #Reasoning #Pre-training

Education

- **Academy for Advanced Interdisciplinary Studies, Peking University** *09.2022 - 06.2025*
  Master of Computer Science, supervised by Prof. Xiaojun Wan GPA: 3.81/4.00
- **School of Electronics Engineering and Computer Science, Peking University** *09.2018 - 06.2022*
  Bachelor of Computer Science GPA: 3.64/4.00

Selected Publications

Preprints:

[1] **ContraSolver: Self-Alignment of Language Models by Resolving Internal Preference Contradictions**
ARR under review [link]
Xu Zhang*, **Xunjian Yin***, Xiaojun Wan.

Conference paper:

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

[2] **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

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

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

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

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

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

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

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

[10] **Evaluating Self-Generated Documents for Enhancing Retrieval-Augmented Generation with LLMs**
NAACL findings 2025 [link]
Jiatao Li, Xinyu Hu, **Xunjian Yin** and 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 Pre-training, Evaluation, Analysis and Applications
    - We trained reverse models in the 2B-7B parameter range from scratch, using a 500 billion token dataset and the last-token prediction approach. We evaluated and analyzed these reverse models with the aim of leveraging them to assist forward models in reasoning, prompt design, and potential adversarial attacks on forward models.
  - **Project 2**: Gödel Agent: A Self-Referential Agents Framework for Recursively Self-Improvement
    - We developed a self-referential agent framework, Gödel Agent, which is capable of reading and modifying its own logic and code. This agent can optimize itself based on feedback from its environment.

- **Microsoft Research Asia (NLC Group)** *02.2022 - 08.2022*
  ***Advisor**: Dr. Shuming Ma* ***Role**: Research Intern* Beijing, China
  - **Project:** Pre-training with Curriculum Learning
    - We investigated the application of curriculum learning during pretraining, and discovered that starting with simple data before progressing to more complex data significantly accelerates the convergence of the pretraining process.
- **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 conducted an analysis of sequence-to-sequence models and methods in data-to-text generation tasks, finding that the most advanced models do not always yield the best performance. However, in general, larger models tend to perform better.
  - **Project 2:** Enhancing Language Models with kNN for Grammar Error Correction
    - In the grammar error correction task, We designed a k-nearest neighbor retrieval algorithm to augment the language model. We proposed incorporating robust information derived from character phonetics and shapes to improve retrieval accuracy, which helps mitigate the interference caused by erroneous information.
- **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
    - In GEC, We applied multi-task learning to train the decoder. 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 contributed to the creation of a benchmark to assess the reasoning capabilities of models, using middle-school-level mathematical Olympiad problems as the testbed.

Honors and Awards

- **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'23, EMNLP'23, Coling'24, ARR 2024 (ACL, EMNLP, NAACL)
  - ICLR'24, NeurIPS'24, ICLR'25, ICML'25, NeurIPS'25
- **Volunteer:** *2023 - Present*
  - AAAI'24, ACL'24
  - NLPCC'23 Shared Task 8 track chair

Skills

- **Professional Skills:** Large Language Model pre-training, fine-tuning, alignment, prompt engineering
- **Languages:** English, Chinese
- **Interests:** squash, badminton, swimming, skiing