Xiaocong Yang

PhD Student, CS @ UIUC | Founder, AI Interpretability @ Illinois

Personal Website | Blog Posts | Google Scholar | LinkedIn | xy51@illinois.edu

Education

Ph.D. in Computer Science | University of Illinois Urbana-Champaign **2025.8 - 2028.5**

- Founder and current leader of *AI Interpretability @ Illinois* research team.

M.S. in Computer Science | University of Illinois Urbana-Champaign **2023.8 - 2025.5**

B.S. in Economics | Tsinghua University **2018.9 - 2023.6**

Publication

7. Xiaocong Yang. LLM2IR: Simple Unsupervised Contrastive Learning Makes Long-Context LLM Great Retriever. MS Thesis [link]

6. Xiaocong Yang, Jiacheng Lin, Ziqi Wang, Chengxiang Zhai. *Learning by Analogy: Enhancing Few-Shot Prompting for Math Word Problem Solving with Computational Graph-Based Retrieval*, preprint [link]

5. Ziyi Chen, **Xiaocong Yang**, Jiacheng Lin, Chenkai Sun, Jie Huang, Kevin Chen-Chuan Chang. *Cascade Speculative Drafting for Even Faster LLM Inference*, **NeurIPS 2024** [link]

4. Xiaocong Yang, James Huang, Wenxuan Zhou, Muhao Chen. *Parameter-efficient Tuning with Special Token Adaptation*, **EACL 2023** [link]

3. Xingcheng Yao, Yanan Zheng, **Xiaocong Yang**, Zhilin Yang. *NLP from Scratch Without Large-Scale Pretraining: A Simple and Efficient Framework,* **ICML 2022** [link]

2. Leixian Shen, Enya Shen, Yuyu Luo, **Xiaocong Yang**, Xuming Hu, Xiongshuai Zhang, Zhiwei Tai, Jianmin Wang. *Towards natural language interfaces for data visualization: A survey*, **TVCG 2022** [link]

1. Hao Zhou, Pei Ke, Zheng Zhang, Yuxian Gu, Yinhe Zheng, Chujie Zheng, Yida Wang, Chen Henry Wu, Hao Sun, **Xiaocong Yang**, Bosi Wen, Xiaoyan Zhu, Minlie Huang, Jie Tang. *Eva: An open-domain chinese dialogue system with large-scale generative pre-training*, preprint [link]

Research Experience

AI Scientist Intern | Intuit Inc. **2025.5-2025.8**

Supervisor: Ruocheng Guo | Staff Research Scientist, Intuit AI Research Mountain View

- Raised RestBench tool-call success by **+5 percentage points** by designing tool description improvement pipeline for agentic LLM flows; scoped and led end-to-end study at Intuit.

Student Researcher | TIMAN Group, University of Illinois Urbana-Champaign **2024.3 - present**

Advisor: Chengxiang Zhai | Donald Biggar Willett Professor, UIUC Urbana-Champaign

- Proposed and led research on *Learning by Analogy*, a RAG system in which the retriever is trained to retrieve examples with similar computational graphs to enhance few-shot prompting by providing the LLM with the correct reasoning path. Our approach outperforms the baseline by **+6.7 percentage points**.
- See full ongoing project list here.

Student Researcher | FORWARD Data Lab, University of Illinois Urbana-Champaign **2023.8 - 2024.1**

Advisor: Kevin Chenchuan Chang | Professor, UIUC Urbana-Champaign

- Proposed *Cascade Speculative Drafting*, where we vertically and horizontally cascade multiple draft models of different size. Our approach is mathematically proved to be more efficient than standard speculative decoding and achieves up to **72%** additional speedup in the experiments.
- Paper is published at NeurIPS 2024.

Visiting Researcher | LUKA Group, University of Southern California **2022.4 - 2022.9**

Advisor: Muhao Chen | Assistant Professor, Information Science Institute, USC Los Angeles

- Proposed and led research on *PASTA*, a model finetuning method that is more efficient than all of existing baselines which tunes only **0.029%** of total parameters to match the performance of full finetuning.
- Paper is published at EACL 2023.

Student Researcher | IIIS, Tsinghua University **2021.7 - 2023.1**
Advisor: Zhilin Yang | Assistant Professor, Institute for Interdisciplinary Information Sciences, Tsinghua University | Founder & CEO, Moonshot AI Beijing

- Contributed to *Task-driven Language Modeling*, proposing a task-specific pretraining framework that reduced by **100x** the total training cost and size of corpus and democratizing the development of large LLMs.
- Analyzed the functionality of "vertical attention heads" in LLMs.
- Paper is published at ICML 2022.

Student Researcher | Conversational AI (CoAI) Group **2021.1 - 2021.6**
Advisor: Huang Minlie | Associate Professor | Department of Computer Science and Technology, Tsinghua University Beijing

- Contributed to the development of the open-domain dialogue system *EVA* and the corresponding paper.
- Designed the interface and client-server communication framework for the deployment of the dialog system, tested rule-based and model-based methods to filter out toxic content in the generated text.

Teaching Experience

- **Instructor** | CS 591 Biologically Plausible AI, UIUC **Fall 2025**
  - Section 1: What is Interpretability, and Why It Matters [Recording] [Slides]
  - Section 2: Neuron-Level Interpretability: Biologically Plausible Neurons [Recording] [Slides]
  - Section 3: Model-Level Interpretability: Transformer Circuits & Modularity [Recording A] [Recording B] [Slides]
  - Section 4: Behaviour-Level Interpretability: Test-Time Compute, Neuro-Symbolic AI, Competitive Market Based Learning [Recording] [Slides]