# Xiaoman Pan

Contact shyowmanpan@gmail.com

## *Research Interests*

My research focuses on machine learning, natural language processing, and large language models. I am currently dedicated to developing next-generation foundational models to support Amazon Stores businesses.

## *Education*

**University of Illinois at Urbana-Champaign (UIUC)** Urbana, IL
*Ph.D., Computer Science* *Dec. 2020*
**Advisor**: Prof. Heng Ji

**Rensselaer Polytechnic Institute (RPI)** Troy, NY
*B.S., Computer Science* *Dec. 2014*
*Minor in Psychology* *GPA 3.93/4.0*

## *Work Experience*

**Amazon, Stores Foundational AI** Seattle, WA
*Senior Applied Scientist* *Sep. 2024 - Present*

**Tencent AI Lab** Bellevue, WA
*Senior Researcher* *Jan. 2021 - Sep. 2024*

**Tencent AI Lab** Bellevue, WA
*Research Intern* *Feb. 2019 - May 2019*

**Facebook Research - Applied Machine Learning** Menlo Park, CA
*Research Intern* *Feb. 2017 - May 2017*

## *Selected Publications*

[35] Hongming Zhang, **Xiaoman Pan**, Hongwei Wang, Kaixin Ma, Wenhao Yu, Dong Yu. Cognitive Kernel: An Open-source Agent System towards Generalist Autopilots. On ArXiv. 2024.

[34] Ruixin Hong, Hongming Zhang, **Xiaoman Pan**, Dong Yu, Changshui Zhang. Abstraction-of-Thought Makes Language Models Better Reasoners. On ArXiv. 2024.

[33] Xinran Zhao, Hongming Zhang, **Xiaoman Pan**, Wenlin Yao, Dong Yu, Tongshuang Wu, Jianshu Chen. Fact-and-Reflection (FaR) Improves Confidence Calibration of Large Language Models. In Findings of the Association for Computational Linguistics ACL 2024.

[32] Rui Yang*, **Xiaoman Pan***, Feng Luo*, Shuang Qiu*, Han Zhong, Dong Yu, Jianshu Chen. Rewards-in-Context: Multi-objective Alignment of Foundation Models with Dynamic Preference Adjustment. Proc. International Conference on Machine Learning (ICML) 2024.

[31] Wenhao Yu, Hongming Zhang, **Xiaoman Pan**, Kaixin Ma, Hongwei Wang, Dong Yu. Chain-of-note: Enhancing robustness in retrieval-augmented language models. On ArXiv. 2023.

[30] Kaiqiang Song, Xiaoyang Wang, Sangwoo Cho, **Xiaoman Pan**, Dong Yu. Zebra: Extending Context Window with Layerwise Grouped Local-Global Attention. On ArXiv. 2023.

[29] Xuansheng Wu, Wenlin Yao, Jianshu Chen, **Xiaoman Pan**, Xiaoyang Wang, Ninghao Liu, Dong Yu. From language modeling to instruction following: Understanding the behavior shift in llms after instruction tuning. Proc. the 2024 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2024).

[28] Kaixin Ma, Hongming Zhang, Hongwei Wang, **Xiaoman Pan**, Dong Yu. Laser: Llm agent with state-space exploration for web navigation. On ArXiv. 2023.

[27] Jiaao Chen, **Xiaoman Pan**, Dian Yu, Kaiqiang Song, Xiaoyang Wang, Dong Yu, Jianshu Chen. Skills-in-context prompting: Unlocking compositionality in large language models. Proc. the 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP2024).

[26] Zhenwen Liang, Dian Yu, **Xiaoman Pan**, Wenlin Yao, Qingkai Zeng, Xiangliang Zhang, Dong Yu. Mint: Boosting generalization in mathematical reasoning via multi-view fine-tuning. Proc. the 2024 Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING 2024).

[25] Xinran Zhao, Hongming Zhang, **Xiaoman Pan**, Wenlin Yao, Dong Yu, Jianshu Chen. Thrust: Adaptively Propels Large Language Models with External Knowledge. Advances in Neural Information Processing Systems. 2023.

[24] Linfeng Song, Ante Wang, **Xiaoman Pan**, Hongming Zhang, Dian Yu, Lifeng Jin, Haitao Mi, Jinsong Su, Yue Zhang, Dong Yu. OpenFact: Factuality Enhanced Open Knowledge Extraction. Transactions of the Association for Computational Linguistics. 2023.

[23] Keming Lu, **Xiaoman Pan**, Kaiqiang Song, Hongming Zhang, Dong Yu, Jianshu Chen. PIVOINE: Instruction Tuning for Open-world Entity Profiling. In Findings of the Association for Computational Linguistics: EMNLP 2023.

[22] Wenlin Yao, Lifeng Jin, Hongming Zhang, **Xiaoman Pan**, Kaiqiang Song, Dian Yu, Dong Yu, Jianshu Chen. How do Words Contribute to Sentence Semantics? Revisiting Sentence Embeddings with a Perturbation Method. Proc. the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2023).

[21] **Xiaoman Pan**, Wenlin Yao, Hongming Zhang, Dian Yu, Dong Yu, Jianshu Chen. Knowledge-in-Context: Towards Knowledgeable Semi-Parametric Language Models. Proc. International Conference on Learning Representations (ICLR), 2023.

[20] Xianjun Yang, Kaiqiang Song, Sangwoo Cho, Xiaoyang Wang, **Xiaoman Pan**, Linda Petzold, Dong Yu. OASum: Large-Scale Open Domain Aspect-based Summarization. In Findings of the Association for Computational Linguistics: ACL 2023.

[19] Zhenhailong Wang, **Xiaoman Pan**, Dian Yu, Dong Yu, Jianshu Chen, Heng Ji. Zemi: Learning Zero-Shot Semi-Parametric Language Models from Multiple Tasks. Proc. the 61th Annual Meeting of the Association for Computational Linguistics Findings (ACL2023).

[18] Pei Chen, Wenlin Yao, Hongming Zhang, **Xiaoman Pan**, Dian Yu, Dong Yu, Jianshu Chen. ZeroKBC: A Comprehensive Benchmark for Zero-Shot Knowledge Base Completion. Proc. The 22nd IEEE International Conference on Data Mining (ICDM).

[17] Xiang Yue, **Xiaoman Pan**, Wenlin Yao, Dian Yu, Dong Yu, Jianshu Chen. C-MORE: Pretraining to Answer Open-Domain Questions by Consulting Millions of References. Proc. the 60th Annual Meeting of the Association for Computational Linguistics (ACL2022).

[16] Wenlin Yao, **Xiaoman Pan**, Lifeng Jin, Jianshu Chen, Dian Yu, Dong Yu. Connect-the-Dots: Bridging Semantics between Words and Definitions via Aligning Word Sense Inventories. Proc. The 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP2021).

[15] Haoyang Wen, Ying Lin, Tuan Lai, **Xiaoman Pan**, Sha Li, Xudong Lin, Ben Zhou, Manling Li, Haoyu Wang, Hongming Zhang, Xiaodong Yu, Alexander Dong, Zhenhailong Wang, Yi Fung, Piyush Mishra,

Qing Lyu, Dídac Surís, Brian Chen, Susan Windisch Brown, Martha Palmer, Chris Callison-Burch, Carl Vondrick, Jiawei Han, Dan Roth, Shih-Fu Chang, Heng Ji. RESIN: A Dockerized Schema-Guided Cross-document Cross-lingual Cross-media Information Extraction and Event Tracking System. Proc. The 2021 Conference of the North American Chapter of the Association for Computational Linguistics - Human Language Technologies (NAACL-HLT2021) Demo Track.

[14] Manling Li, Alireza Zareian, Ying Lin, **Xiaoman Pan**, Spencer Whitehead, Brian Chen, Bo Wu, Heng Ji, Shih-Fu Chang, Clare Voss, Daniel Napierski and Marjorie Freedman. GAIA: A Fine-grained Multimedia Knowledge Extraction System. Proc. The 58th Annual Meeting of the Association for Computational Linguistics (ACL2020) Demo Track (Best Demo Paper).

[13] **Xiaoman Pan***, Kai Sun*, Dian Yu, Jianshu Chen, Heng Ji, Claire Cardie and Dong Yu. Improving Question Answering with External Knowledge. Proc. EMNLP2019 Workshop on Machine Reading for Question Answering.

[12] **Xiaoman Pan**, Thamme Gowda, Heng Ji, Jonathan May and Scott Miller. Cross-lingual Joint Entity and Word Embedding to Improve Entity Linking and Parallel Sentence Mining. Proc. EMNLP2019 Workshop on Deep Learning for Low-Resource Natural Language Processing.

[11] Qingyun Wang, **Xiaoman Pan**, Lifu Huang, Boliang Zhang, Zhiying Jiang, Heng Ji and Kevin Knight. Describing a Knowledge Base. Proc. The 11th International Conference on Natural Language Generation.

[10] Boliang Zhang, Ying Lin, **Xiaoman Pan**, Di Lu, Jonathan May, Kevin Knight and Heng Ji. ELISA-EDL: A Cross-lingual Entity Extraction, Linking and Localization System. Proc. The 16th Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT2018) Demo Track.

[9] **Xiaoman Pan**, Boliang Zhang, Jonathan May, Joel Nothman, Kevin Knight and Heng Ji. Cross-lingual Name Tagging and Linking for 282 Languages. Proc. The 55th Annual Meeting of the Association for Computational Linguistics (ACL2017).

[8] Lifu Huang, Jonathan May, **Xiaoman Pan**, Heng Ji, Xiang Ren, Jiawei Han, Lin Zhao and James Hendler. Liberal Entity Extraction: Rapid Construction of Fine-Grained Entity Typing Systems. Big Data, Mar 2017, 5(1): 19-31.

[7] Dongxu Zhang, Boliang Zhang, **Xiaoman Pan**, Xiaocheng Feng, Heng Ji, Weiran Xu. 2016. Bitext Name Tagging for Cross-lingual Entity Annotation Projection. Proc. The 26th International Conference on Computational Linguistics (COLING 2016).

[6] Ellie Pavlick, Heng Ji, **Xiaoman Pan**, Chris Callison-Burch. 2016. The Gun Violence Database: A new task and data set for NLP. Proc. Conference on Empirical Methods in Natural Language Processing (EMNLP 2016).

[5] Ying Lin, **Xiaoman Pan**, Aliya Deri, Heng Ji, Kevin Knight. 2016. Leveraging Entity Linking and Related Language Projection to Improve Name Transliteration. Proc. ACL2016 Workshop on Named Entities.

[4] Di Lu, **Xiaoman Pan**, Nima Pourdamghani, Shih-Fu Chang, Heng Ji, Kevin Knight. 2016. A Multi-media Approach to Cross-lingual Entity Knowledge Transfer. Proc. The 54th Annual Meeting of the Association for Computational Linguistics (ACL 2016).

[3] Chuan Wang, Sameer S Pradhan, **Xiaoman Pan**, Heng Ji, Nianwen Xue. 2016. CAMR at SemEval-2016 Task 8: An Extended Transition-based AMR Parser. Proc. NAACL-HLT 2016 Workshop on Semantic Evaluation (SemEval-2016).

[2] Boliang Zhang, **Xiaoman Pan**, Tianlu Wang, Ashish Vaswani, Heng Ji, Kevin Knight, Daniel Marcu. 2016. Name Tagging for Low-resource Incident Languages based on Expectation-driven Learning. Proc. The 2016 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL-HLT 2016).

[1] **Xiaoman Pan**, Taylor Cassidy, Ulf Hermjakob, Heng Ji, Kevin Knight. 2015. Entity Linking with Abstract Meaning Representation. Proc. The 2015 Conference of the North American Chapter of the Association for Computational Linguistics – Human Language Technologies (NAACL HLT 2015).

## *Services*

**Program Committee**: ACL (2017-23), NAACL (2016, 2018, 2022), EMNLP (2018-19), COLING (2018, 2020), LREC (2020, 2022)

## *Professional Skills*

Programming Languages/Tools: Python, PyTorch, MongoDB, Docker, Emacs