

Yuxuan Jiang

✉ yuxuanj1@umbc.edu 🌐 yuxuanjiang1.github.io 📞 +1-551-998-5415

Education

University of Maryland, Baltimore County, Baltimore, MD Sep. 2023 – Present
Ph.D. Student in Computer Science Advisor: Prof. Francis Ferraro
Research Focus: Large Language Models, Tool-Augmented Reasoning, NLP Efficiency

New York University, New York, NY Jan. 2021 – Dec. 2022
Master of Science in Computer Science GPA: 3.6/4.0

Beijing Language and Culture University, Beijing, China Sep. 2016 – Jul. 2020
Bachelor of Information Systems GPA: 3.2/4.0

Experience

Xiaomi, Beijing, China Jul. 2020 – Dec. 2020
Machine Learning Engineer Intern

- Built the first large-scale Chinese commonsense knowledge graph for multi-turn dialogue systems, enhancing contextual response diversity in production chatbot pipelines.
- Translated and localized the ATOMIC dataset into Chinese; curated a 200K+ multi-turn daily conversation corpus for dialogue training and evaluation.
- Collaborated cross-functionally to integrate commonsense reasoning into conversational AI systems used in commercial voice assistants.
- **Tech Stack:** Python, PyTorch, Transformers, Neo4j, RESTful APIs.

National Language Resource Monitoring Center, Beijing, China Jan. 2019 – Jun. 2020
Research Assistant

- Reduced false-touch rate of a 26-key mobile input method by redefining keyboard boundaries, improving user typing accuracy by 15% in field tests.
- Developed a multi-task span-prediction model combining syntactic and semantic supervision, improving SRL F1 by 4.6%.
- Authored a technical report on personalized cloud-based text error correction integrating contextual modeling of user input behaviors.
- **Tech Stack:** TensorFlow, BERT, Flask, MySQL, Cloud-based APIs.

Publications

1. “From Generation to Judgment: Opportunities and Challenges of LLM-as-a-Judge”, *EMNLP 2025 Main Conference*.
2. “Memorization Over Reasoning? Exposing and Mitigating Verbatim Memorization in Large Language Models’ Character Understanding Evaluation”, under review at *ACL 2026* (arXiv).
3. “DRP: Distilled Reasoning Pruning with Skill-aware Step Decomposition for Efficient Large Reasoning Models”, under review at *ACL 2026* (arXiv).
4. “Learning How to Use Tools, Not Just When: Pattern-Aware Tool-Integrated Reasoning”, MATH-AI Workshop, *Neurips 2025* (arXiv).

Programming Skills

- **Languages:** Python, C++, SQL
- **Frameworks:** PyTorch, Hugging Face Transformers, TensorFlow
- **Tools:** GCP, AWS, Docker, Git, Linux