

Tianyue Ou

tianyueo@andrew.cmu.edu

EDUCATION

Carnegie Mellon University Pittsburgh, PA
Master of Science 2023 - 2025
Courses: Advanced Natural Language Processing, Multimodal Machine Learning, Large Language Model Systems,
Multilingual Natural Language Processing, Neural Code Generation

Johns Hopkins University Baltimore, MD
Bachelor of Science, Computer Science 2017 - 2021
Courses: Machine Learning: Deep Learning, Machine Learning, Natural Language Processing

RESEARCH INTEREST

LLM-Based Agent, Synthetic Data, Multi-Agent Reasoning, Benchmarking LLM

RESEARCH(* denotes equal contribution)

Synatra: Turning indirect knowledge into direct demonstrations for digital agents at scale[[link](#)] NeurIPS 2024
[Tianyue Ou](#), [Frank F Xu](#), [Aman Madaan](#), [Jiarui Liu](#), [Robert Lo](#), [Abishek Sridhar](#), [Sudipta Sengupta](#), [Dan Roth](#), [Graham Neubig](#), [Shuyan Zhou](#).

- We developed state of the art open-sourced web agent by training with synthetic data. Synatra, the data synthesis approach, leverages indirect knowledge from tutorials and web snapshots to construct agent trajectories. Agents trained with Synatra-generated trajectories outperform those trained with human annotated ones.

Harnessing Webpage UIs For Text Rich Visual Understanding[[link](#)] Under Review 2024
[Junpeng Liu](#), [Tianyue Ou*](#), [Yifan Song*](#), [Yuxiao Qu*](#), [Wai Lam](#), [Chenyan Xiong](#), [Wenhu Chen](#), [Graham Neubig](#), [Xiang Yue](#)

- We created a training dataset that provide models with state of the art visual understanding, grounding, and agent performance. The dataset, MultiUI, is created by combining visual information in webpages with structural information in accessibility trees.

An In-depth Look at Gemini's Language Abilities[[link](#)] Preprint 2023
[Syeda Nahida Akter*](#), [Zichun Yu*](#), [Aashiq Muhamed*](#), [Tianyue Ou*](#), [Alex Bäuerle](#), [Ángel Alexander Cabrera](#), [Krish Dholakia](#), [Chenyan Xiong](#), [Graham Neubig](#)

- We analyzed Gemini's performance in web agents, reasoning, math, code generation, machine translation, and knowledge-based QA.

CowPilot: A Framework for Autonomous and Human-Agent Collaborative Web Navigation Under Review 2024
[link](#)

[Faria Huq](#), [Zora Zhiruo Wang](#), [Frank F. Xu](#), [Tianyue Ou](#), [Shuyan Zhou](#), [Jeffrey P. Bigham](#), [Graham Neubig](#)

- We made a copilot tool that allows users to seamlessly collaborate with web agents when they are browsing. We analyzed through experiments to find that human-agent collaborative mode achieves the highest success rate in task completion.

Webarena: A realistic web environment for building autonomous agents[[link](#)] ICLR 2024
[Shuyan Zhou](#), [Frank F Xu](#), [Hao Zhu](#), [Xuhui Zhou](#), [Robert Lo](#), [Abishek Sridhar](#), [Xianyi Cheng](#), [Tianyue Ou](#), [Yonatan Bisk](#), [Daniel Fried](#), [Uri Alon](#), [Graham Neubig](#)

- We built an interactive web agent benchmark that focuses on evaluating the functional correctness of task completions. We built a diverse set of tasks including web shopping, gitlab collaboration, map usage, forum discussion, and so on.

Toward Global AI Inclusivity: A Large-Scale Multilingual Terminology Dataset Under Review 2024
[Jiarui Liu](#), [Iman Ouzzani](#), [Wenkai Li](#), [Lechen Zhang](#), [Tianyue Ou](#), [Houda Bouamor](#), [Zhijing Jin](#), [Mona T. Diab](#)

- We created GIST, a large-scale multilingual AI terminology dataset of 5K terms extracted from top AI conference papers. The terms were translated into Arabic, Chinese, French, Japanese, and Russian.

EXPERIENCE

Research Assistant

2023 - Current

NeuLab, CMU

- Advised by Prof. Graham Neubig and Shuyan Zhou.
- Developed web and GUI agents through building synthetic data. Focusing on leveraging indirect knowledge from existing information.
- Building multimodal reasoning benchmark.
- Involved in building agent benchmark and evaluating existing model in agent settings.

Research Assistant

2024 - Current

Daniel Fried Lab, CMU

- Advised by Prof. Daniel Fried.
- Focus on reasoning with multi-agent, studying the impact of multi-agent for planning in complex scenarios.

Machine Learning Engineer

2022 - 2023

Meta

- Worked on music recommendation for facebook stories and reels.
- Improved consumption and creation metrics with user interests embeddings, personalized page rank embeddings, etc.

SKILLS

Proficient in Python, Java, C/C++, OCaml

Experienced with popular packages such as Pytorch, Llamafactory, Deepspeed, Autogen, TGI, vLLM

Experienced in backend tools such as Dataswarm, Hive, Spark, SQL