Yu Ying Chiu (Kelly)

kellycyy@uw.edu | (+1) 206 227 7140 | GScholar: Yu Ying Chiu | Linkedin: kellycyy | Github: @kellycyy

Education

University of Washington, MS in Computational Linguistics (NLP) Sep 2022 – Dec 2024

- Natural Language Processing, Language Models & Reasoning, Artificial Intelligence, GPA: 4.0/4.0

University of Hong Kong, BS in Decision Analytics (Stat./Comp. Sci.)& Psychology Sep 2017 – Jun 2022

- Statistical Machine Learning & Computational Social Psychology
- **Thesis:** Understanding Human Decision-Making in Stock Markets through Social Media Sentiment Analysis

Work Experience

Allen Institute for Artificial Intelligence, Research Collaborator May 2023 – Now

- Evaluated and benchmarked cultural knowledge of LLMs, advised by Dr. Bill Yuchen Lin and Prof. Yejin Choi.
- Built human-in-the-loop data collection platform for collecting 2000^+ samples from 45^+ countries.

University of Washington Allen School of Computer Science, Research Assistant May 2023 – Now

- Designed evaluation framework and finetuned models for assessing LLM therapists behaviors during psychotherapy (e.g. reflecting upon client needs, normalizing expectations), advised by Prof. Tim Althoff.
- Investigated the value preference of models in relation to psychological, sociological and philosophical theories, to provide insights on model alignment (e.g. effectiveness of AI Constitutions), advised by Prof. Yejin Choi.

Hong Kong Monetary Authority, Data Scientist Sep 2020 – Apr 2021

- Designed and implemented analysis tool on predicting stock market with real-time social media posts using time-series ML algorithm and BERT model for more than 50^+ researchers.

Publications

- **Yu Ying Chiu**, Liwei Jiang, Bill Yuchen Lin, Chan Young Park, Shuyue Stella Li, Sahithya Ravi, Mehar Bhatia, Maria Antoniak, Yulia Tsvetkov, Vered Shwartz, Yejin Choi. 2024. CulturalBench: a Robust, Diverse and Challenging Benchmark on Measuring the (Lack of) Cultural Knowledge of LLMs. *Under Conference Review.* [**Paper | Data | Leaderboard**]
- **Yu Ying Chiu**, Liwei Jiang, Yejin Choi. 2024. DailyDilemmas: Revealing Value Preferences of LLMs with Quandaries of Daily Life. *Under Conference Review.* [**Paper | Code | Data**]
- Wenting Zhao, Tanya Goyal, **Yu Ying Chiu**, Liwei Jiang, Benjamin Newman, Abhilasha Ravichander, Khyathi Chandu, Ronan Le Bras, Claire Cardie, Yuntian Deng, Yejin Choi. 2024. WildHallucinations: Evaluating Long-form Factuality in LLMs with Real-World Entity Queries. *Under Conference Review.* [**Paper**]
- **Yu Ying Chiu***, Ashish Sharma*, Inna Wanyin Lin, Tim Althoff. 2024. A Computational Framework for Behavioral Assessment of LLM Therapists. *Under Journal Review.* [**Paper | Code+Data**]
- Zhilin Wang*, **Yu Ying Chiu***, Yu Cheng Chiu. 2023. Humanoid Agents: Platform for Simulating Human-like Generative Agents. *Accepted in EMNLP System Demo 2023.* [**Paper | Code | Demo**]
- **Yu Ying Chiu**, Liwei Jiang, Maria Antoniak, Chan Young Park, Shuyue Stella Li, Mehar Bhatia, Sahithya Ravi, Yulia Tsvetkov, Vered Shwartz, Yejin Choi. 2024. CulturalTeaming: AI-Assisted Interactive Red-Teaming for Challenging LLMs'(Lack of) Multicultural Knowledge. *Preprint.* [**Paper**]
- Abhinav Patil, Jaap Jumelet, **Yu Ying Chiu**, Andy Lapastora, Peter Shen, Lexie Wang, Clevis Willrich, Shane Steinert-Threlkeld. 2023. Filtered Corpus Training (FiCT) Shows that Language Models can Generalize from Indirect Evidence. *Accepted in TACL.* [**Paper**]

Technical Skills

Languages: Python, SQL, JavaScript, HTML/CSS, R, MATLAB
NLP/ML skills: Pytorch, Scikit-learn, NLTK, SpaCy, Flair,Statsmodels