Yueru Yan

Bloomington, Indiana | yueryan@iu.edu | 626-998-1352

Education

Indiana University, Bloomington, *PhD in Information Science & Minor in Computer Science* Sept 2023–May 2027(expected)

Beijing Normal University, *M.S. in Applied Statistics* Sept 2021–June 2023

Beijing Foreign Studies University, *B.S. in Journalism* Sept 2015–June 2019

Technical Skills

- **Machine Learning & NLP (Proficient):** supervised/unsupervised learning, sequence modeling (RNN, LSTM, Transformers), reinforcement learning, LLM fine-tuning, knowledge distillation, model evaluation (accuracy, F1, ROC-AUC).
- **Statistical Analysis (Expert):** regression modeling (linear, logistic, mixed-effects), generalized linear/multivariate models, Bayesian inference, hypothesis testing, ANOVA, time-series analysis, non-parametric methods, experimental/study design.
- **Data Mining & Text Analytics:** large-scale data collection, text analysis, topic modeling, sentiment/network analysis.
- **Programming & Tools:** Python, SQL; **Libraries**: PyTorch, Hugging Face Transformers, scikit-learn, TensorFlow, spaCy, NLTK, NumPy, pandas, Matplotlib, Seaborn. **Systems**: Git/GitHub, Linux/Unix shell scripting, Docker, Conda.

Work Experience

Samsung Research Center (Language Understanding Lab) Beijing, China

Data and Machine Learning Engineering (Internship) Dec 2021 – Jan 2023

- Contributed to a $400K funded project & Published two papers & Won the Best Intern Award.
- Extracted multi-modal features and processed sequential data from participants' smartphones, cameras, and sensors to detect mental health issues; Developed and applied sequential machine learning models, including CNN, LSTM, and a custom-designed DEP-CASER model, for depression detection, achieving a state-of-the-art accuracy of 0.83.

Research Experience

Implicit Profiling in LLMs Jan 2025 - Present

- Designed a Bayesian evaluation of contextual attribute inference in LLMs by computing teacher-forced log-likelihoods for candidate responses, contrasting contexts via log-likelihoods, and converting to posteriors (softmax) to measure personalization for attributes like gender, race and political leaning.
- Quantified how context-window length affects inference strength using window-size posteriors.

Large-Scale Multi-Platform LLM Interaction Dataset April 2023 – Present

- Constructed a large-scale, multi-platform conversational corpus (142K+ conversations) that aggregates native interaction features, addressing a critical gap in evaluating model reliability and user interaction within complex, multi-turn scenarios.
- Conducted conversation completeness analysis, temporal analysis and source analysis, demonstrating substantial variation in user intent resolution that aligns with each platform's functional emphasis.

AINotator: LLM Annotation Framework June 2024 – Present

- Co-developed an open-source framework for large-scale annotation using LLMs, integrating multiple APIs (OpenAI, Anthropic, etc.) with automated pipelines for JSON parsing, debugging, and evaluation.
- Enabled annotation beyond surface-level text by capturing contextual and implied meanings in sentences, supporting nuanced communicative signals like speech acts and politeness.

TikTok Comment Personalization & Algorithm Auditing Sept 2024 - Sept 2025

- Designed and built a large-scale TikTok political video & comment dataset (4,691 videos, 10M+ comments) and used LLM to analyze how user ideological leanings influence comment ranking under political videos.
- Conducted an audit experiment to see how the comments are personalized to different users and found that personalization exists at TikTok comment section and right leaning accounts tend to see more right-leaning content at the top positions.

Publications

Under Review:

- **Auditing Algorithmic Personalization in TikTok Comment Sections, Yueru Yan**, Siqi Wu, *ICWSM 2026*, Sept 2025
- **Fairness Evaluation of Large Language Models in Academic Library Reference Services,** Haining Wang, Jason Clark, **Yueru Yan**, et al. *Humanities and Social Sciences Communications*, July 2025

Published:

- **Hardship streamers in live crowdfunding, Yueru Yan**, Pnina Fichman, *iConference*, March 2025
- **A CNN Model with Discretized Mobile Features for Depression Detection, Yueru Yan** et al., *IEEE BSN*, Aug 2022
- **Detecting Depression, Anxiety and Mental Stress in One Sequential Model with Multi-task Learning,** Shen Zhang, Mei Tu, **Yueru Yan**, Yimeng Zhuang, Likun Ge, Gaoxia Wei, *HCI International – Late Breaking Papers*, May 2022