

Yinuo (Nancy) Xu
PhD Candidate at University of Michigan, NLP & LLM Researcher
(626)427-5725 · yinuoxu@umich.edu · nancy078.github.io · linkedin.com/nancy-xu-7840491b1

Education

University of Michigan, Ann Arbor

Ph.D. in Information

University of California, Berkeley

2023 – Present

Advisor: David Jurgens

B.A. Data Science & Statistics, 2023

Awards & Leadership

- Organizer, Michigan NLP Reading Group, 2025-2026
 - Organizer, NLP@Michigan Day 2026, 2025-2026
 - Science Communication Fellow (selected as 1 of 20 fellows university-wide), 2023–2024
 - University of Michigan School of Information Merit Fellowship (2023)
-

Publications

- **Xu, Y.**, Derricks, V., Earl, A., & Jurgens, D. (2025). *Modeling annotator disagreement with demographic-aware experts and synthetic perspectives*. Accepted to the Association for Computational Linguistics (ACL 2026 Oral Presentation). <https://aclanthology.org/2026.acl-long.1914/>
 - **Xu, Y.**, & Jurgens, D. (2026). *Beyond consensus: Perspectivist modeling and evaluation of annotator disagreement in NLP*. Accepted at EMNLP 2026. arXiv:2601.09065
 - **Xu, Y.**, Chen, H., Rakshit, S., *et al.*, & Jurgens, D. (2025). *Causally modeling the linguistic and social factors that predict email response*. NAACL 2025. (Equal contribution) <https://aclanthology.org/2025.naacl-long.594/>
 - Borchers, C., **Xu, Y.**, & Pardos, Z. A. (2025). *Workload Overload? Late Enrollment Leads to Course Dropout*. Journal of Educational Data Mining, 17(1), 126–156. <https://doi.org/10.5281/zenodo.14907388>
 - Borchers, C., **Xu, Y.**, & Pardos, Z. A. (2024). *Are you an early dropper or late shopper? Mining enrollment transaction data to study procrastination in higher education*. Proceedings of the 17th International Conference on Educational Data Mining (EDM '24). International Educational Data Mining Society.
 - **Xu, Y.**, & Pardos, Z. A. (2024). *Extracting course similarity signal using subword embeddings*. Learning Analytics and Knowledge Conference (LAK '24). ACM. <https://doi.org/10.1145/3636555.3636903>
 - **Xu, Y.**, & Pardos, Z. A. (2023). *Mining detailed course transaction records for semantic information*. Proceedings of the 16th International Conference on Educational Data Mining (EDM '23). International Educational Data Mining Society. <https://doi.org/10.5281/zenodo.8115794>
-

Relevant Experience

Research & Development Intern, Kitware

May 2026 – Present

Project: *Personalized LLM Steering from Inferred Value Latents*

- Designed an end-to-end pipeline that infers **individual users' latent value profiles from a few survey responses**, with a Bayesian active-learning module adaptively selecting the most informative elicitation questions.
- Built a variational information-bottleneck encoder compressing responses into a low-dimensional value embedding with demographic priors for cold-start, then a hypernetwork **amortizing the latent-to-steering-vector map so new users require no optimization**, gated by posterior uncertainty.
- Evaluate against SOTA steering methods across four value datasets: **+0.22 user-specific accuracy over zero-shot**, outperforming existing steering methods.
- Interpretability analysis recovering a **binding-vs-individualizing moral foundations axis** in the learned steering vectors.

Lead Researcher, University of Michigan School of Information

Aug 2023 – Present

Project 1: *Modeling Annotator Disagreement with Demographic-Aware Experts*

- Designed **Demographic-Aware Mixture-of-Experts**, an interpretable architecture with demographic-aware routing to capture structured variation in annotation behavior & intersectional perspectives.
- Achieved state-of-the-art subgroup-level performance across five benchmark datasets, particularly under data imbalance or sparse subgroup representation
- Evaluated LLM annotation reliability through zero-shot, few-shot, and LoRA-finetuned experiments, measuring alignment between persona-prompted LLM ratings and human judgments.
- Developed data-efficient pipelines blending real and LLM-generated synthetic annotations with alignment-weighted loss; implemented multi-GPU training workflows for model scaling and reproducibility; paper **accepted at ACL 2026** arXiv:2508.02853.

Project 2: *B-HAP: Bayesian Hierarchical Adapter Personas for Situated Judgment Modeling*

- Design a hypernetwork-based framework generating parameter-efficient LoRA adapters from task, data, and annotator-feature interactions, with a three-level Bayesian hierarchy (population, subpopulation, individual) to align model behavior with group norms.
- Leading evaluation across a 140K-instance benchmark spanning normative, pragmatic, and preference-judgment tasks, benchmarked against Jury Learning, Mixture-of-Personas, and sociodemographic prompting baselines.
- Originated the core research idea and served as primary writer on an NSF grant proposal for this project; **recommended for funding** by program reviewers (final award pending).

Lead Researcher, UMich School of Information & Social Psychology

Aug 2023 – Present

Project: Measuring Racial Disparities in Doctor-Patient Conversations

- Built a large-scale **annotation pipeline** to evaluate doctor communication quality using a diverse pool of crowdworkers; designed and deployed a custom annotation interface to ensure demographic balance and inter-rater reliability.
- Developed a **zero-shot LLaMA-70B annotation pipeline** to label doctor behaviors (respect, formality) with prompt-engineered templates, validating alignment with human ratings.
- Designed and implemented a **causal mediation model** linking race → doctor language → respect perception → patient trust, revealing differential behavior-trust pathways across demographic groups.
- Paper in preparation for *PNAS*.

Graduate Research Assistant, UMich School of Information

March 2024

Project: Causal Modeling of Linguistic and Social Factors in Email Response (NAACL 2025)

- Co-led the development of **causal inference models** analyzing linguistic and social predictors of email responsiveness, combining structural equation modeling with LLM-based feature extraction.
- Conducted large-scale **language and behavior modeling** using Transformer-based encoders and mixed-effects regression to uncover latent causal drivers of social reciprocity.
- **Presented poster at NAACL 2025**, paper accepted in main [aclanthology.org/2025.naacl-long.594].

NLP Intern, Evisort (acquired by Workday)

2021

- Developed algorithms to redact sensitive information in legal contracts using the Google DLP API and custom regex/XML pipelines.
- Trained and deployed **LightGBM** and **bi-LSTM NER** models in **TensorFlow** for contract field extraction on a SaaS platform used by enterprise clients.
- Designed **PowerBI dashboards** and **PostgreSQL queries** to analyze document metadata and usage patterns from **AWS** data sources.

Teaching Experience

- **Becoming a Data Scientist (SI 568)**, Graduate Student Instructor, University of Michigan, Winter 2026
- **Applied Machine Learning (SI 670)**, Graduate Student Instructor, University of Michigan, Fall 2025
- **Natural Language Processing (Info 159/259)**, Teaching Assistant, UC Berkeley, Spring 2023
- **Principles and Techniques of Data Science (Data 100/200)**, Undergraduate Student Instructor, UC Berkeley, Fall 2022
- **Principles and Techniques of Data Science (Data 100/200)**, Undergraduate Student Instructor, UC Berkeley, Summer 2022
- **Technology and Social Impact (EECS 198)**, Lead Facilitator, UC Berkeley, Spring 2023

Professional Service

- Program Committee Member, International Conference on Computational Social Science (IC2S2) 2026
- Program Committee Member, Workshop on NLP and Computational Social Science (at ACL) 2026

Technical Skills Python · PyTorch · Transformers · vLLM · Optuna · ModernBERT · LoRA Fine-Tuning · Mixture-of-Experts · Distributed / Multi-GPU Training (PyTorch DDP) · Reinforcement Learning · Synthetic Data Generation · Causal Inference · Predictive & Statistical Modeling · LLM Evaluation & Alignment · Instruction-Tuning · Data Efficiency · Prompt Engineering · Data Pipeline Design · Bayesian Modeling
