

MINGQIAN ZHENG

Ann Arbor, MI elisazmq@umich.edu

EDUCATION

University of Michigan

Sep.2022 - May.2024

Master in Survey and Data Science (Data Science Track)

Cumulative GPA:4.0/4.0

Coursework: Natural Language Processing, Information Retrieval, Statistical Methods and Machine Learning, Machine Learning for Social Science

Advisor: Prof. David Jurgens

New York University Shanghai

Sep.2018 - May.2022

Bachelor of Science in Mathematics; Bachelor of Science in Data Science (Computer Science Track)

Cumulative GPA:3.7/4.0

Coursework: Machine Learning, Algorithms, Data Structures, Partial Differential Equations, Information Visualization, Databases, Numerical Analysis, Math Modeling, Probability and Statistics

Thesis: User Intention Detection and Evaluation for Recommendation Systems (*Advisors: Prof. Hongyi Wen and Prof. Oliver Marin*)

PUBLICATION

Is “A Helpful Assistant” the Best Role for Large Language Models? A Systematic Evaluation of Social Roles in System Prompts, arXiv:2311.10054, Submitted to NAACL 2024

Mingqian Zheng, Jiaxin Pei, David Jurgens

AWARDS

Rackham Graduate Student Research Grant at University of Michigan

Aug.2023

University Honors Scholar at New York University

May.2022

NYU Shanghai 2020 Recognition Award (Top 0.5%) academic merit and community contributions

Aug.2020

RESEARCH EXPERIENCE

The Blablalab

May.2023 - now

Advisor: Prof. David Jurgens at University of Michigan

- Built a pipeline for systematic evaluation of social roles in Large Language Models' system prompts
- Created a list of 162 roles covering 6 types of interpersonal relationships and 8 types of occupations, and showed that adding interpersonal roles in prompts consistently improves the models' performance over a range of questions through extensive analysis of 3 popular LLMs and 2457 questions
- Conceived and implemented multiple role-prediction methodologies, optimizing for the most effective role assignment in response to specific queries by evaluating frequency, similarity, and perplexity parameters

Global Health and Population Project on Access to Care for Cardiometabolic Diseases (HPACC) Nov.2022 - now

Advisor: Prof. David Flood at University of Michigan

- Independently conducted data harmonization following the World Health Organization (WHO) STEPwise approach to noncommunicable disease risk factor surveillance (STEPS) using Stata
- Assisted in an ongoing research project on predicting diabetes prevalence and progress toward four WHO Diabetes Targets at the country level using a Bayesian statistical modeling approach
- Participated in monthly research meetings with the research team at Heidelberg University

Investigating User Profiles on Reddit: A Multi-Label Classification Study

Mar.2023 - May.2023

Advisor: David Jurgens at University of Michigan - Ann Arbor

- Developed a multi-label classifier using RedDust dataset on Reddit, predicting user attributes (profession, gender, hobbies) through sentence transformer-based text embeddings

- Executed K-means algorithm for community detection on Reddit, analyzing qualitative aspects for insightful community dynamics

Child and Adolescent Data Lab

Oct.2022 - May.2023

Advisor: Prof. Joseph Ryan at University of Michigan

- Conducted thorough data processing and analysis of surveys distributed by the State Court Administrative Offices on the topics of child welfare and juvenile justice systems

User Intention Detection and Evaluation for Recommendation Systems

Feb.2022 - May.2022

Advisors: Prof. Hongyi Wen and Prof. Oliver Marin at NYU Shanghai

- Proposed an intention-aware recall metric along with a statistical mechanics of user intention extraction based on the MovieLens 1M dataset

- Verified the reliability of the metric in evaluating recommendation models by comparing the scores and performances of four DNN and RNN models

INTERNSHIP EXPERIENCE

Natural Language Processing Research Intern(Core R&D Group) at iFlytek

Jul.2021 - Oct.2021

- Applied Recurrent Neural Networks (RNN), sequence labelling and supervised classification methods to implement word-level language identification in a code-switching dataset containing English and Spanish

- Trained the model to achieve accuracy 86% based on the cleaned dataset of 3332 English-Spanish multilingual tweets and the model was used as the baseline model for the future product

Overseas Marketing Intern at HappieWatch

Jun.2021 - Jul.2021

- Conducted in-depth analysis of Instagram and YouTube influencer marketing data, leading to the successful negotiation and signing of contracts with three key opinion leaders (KOLs) for digital marketing campaigns.

- Efficiently supported digital marketing strategies through social media, analyzing Shopify and Google Ads data to deliver insightful weekly reports to stakeholders

EXTRACURRICULAR/LEADERSHIP EXPERIENCE

NYU Shanghai Youth League

Sep.2019 - May.2022

Vice Secretary (Class of 2022)

- Organized and executed more than 10 school-level activities promoting cross-cultural exchange, building a platform for service and community engagement for the entire student body

NYU Shanghai student publication On Century Avenue

Sep.2019 - May.2020

Association Initiator and Director of News and Chinese Committee

- Collaborated with a diverse team to establish an independent student publication, ensuring representation of the entire student body

- Served as a key leader in the initial core group of 6, instrumental in expanding the team to over 80 members, significantly enhancing on-campus presence, evidenced by a 500-follower increase on Instagram

- Led the committee in conducting interviews and reporting on-campus issues, effectively bridging communication between university administration and students

Shanghai Students' Federation

Aug.2018 - May.2020

Group Leader of New Media Department

- Coordinated 20 members to write more than 50 articles on trending issues pertinent to Shanghai college students, encompassing news and survey reports

SKILLS

Programming Languages

Python, R, Stata, Matlab, SQL, HTML/CSS

Programming Libraries

HuggingFace, transformers, Pandas, scikit-learn, NumPy, Matplotlib

Deep Learning Frameworks

PyTorch, TensorFlow, Keras

Statistical and Data Analysis Skills

Regression Analysis, Multilevel Modeling, Causal Inference