Bowen **Yi**

+1 858-776-0900 bowenyi-pierre.github.io bowenyi@umich.edu @BowenYi2003

Research Interests

My research interests lie in **Human-Centered NLP** and **Human-AI Alignment**, with a focus on modeling how personality and cultural differences impact individual behavior across text, speech, and visual modes. I aim to develop AI agents that interpret social and behavioral cues to offer personalized responses, with applications in mental health and education.

Education

Jan. 2023 – Present | **University of Michigan - Ann Arbor** | **Ann Arbor, MI**
B.S. in **Computer Science** | GPA: 3.89/4.0
Coursework: Natural Language Processing, Human-Centered ML, Machine Learning, Computer Vision, Information Retrieval, Human-Centered Software Design and Development

Sep. 2021 – Dec. 2022 | **University of California - San Diego** | **La Jolla, CA**
B.S. in **Mathematics-Computer Science**, Minor in **Sociology** | GPA: 3.83/4.0
Coursework: Sociology of Health Care Issues, Introduction to Sociology, Statistics (in Psychology), Languages and Cultures, Linear Algebra, Calculus, Programming and Object-Oriented Design

Publications

S=In Submission, C=Conference

[C.1] **The Generation Gap: Exploring Age Bias Underlying in the Value Systems of Large Language Models** [🔗]
Siyang Liu and Trish Maturi and Bowen Yi and Siqi Shen and Rada Mihalcea
Conference on Empirical Methods in Natural Language Processing, Miami, USA [**EMNLP 2024**]

[S.4] **Uncovering the Impact of Intervention Messages on Diverse Population Groups**
Bowen Yi, Rada Mihalcea, Fang Yu, Elena Frank, Joan Zhao, Srijan Sen, Maggie Makar
[Working Paper]

[S.3] **Examining Spanish Counseling with MIDAS: a Motivational Interviewing Dataset in Spanish** [🔗]
Aylin Gunal, Bowen Yi, John Piette, Rada Mihalcea, Veronica Perez-Rosas
[Under Review at NAACL 2025]

[S.2] **Real or Robotic? Assessing Whether LLMs Accurately Simulate Qualities of Human Responses in Dialogue** [🔗]
Jonathan Ivey*, Shivani Kumar*, Jiayu Liu*, Hua Shen*, Sushrita Rakshit*, Rohan Raju*, Haotian Zhang*, Aparna Ananthasubramaniam*, Junghwan Kim*, Bowen Yi*, Dustin Wright*, Abraham Israeli*, Anders Giovanni Møller*, Lechen Zhang*, David Jurgens (* = Equal Contribution with Random Order)
[Under Review at NAACL 2025]

[S.1] **Causally Modeling the Linguistic and Social Factors that Predict Email Response** [🔗]
Yinuo Xu*, Sushrita Rakshit*, Aparna Ananthasubramaniam*, Omkar Yadav*, Mingqian Zheng*, Michael Jiang*, Lechen Zhang*, Bowen Yi*, Kenan Alkiek*, Abraham Israeli*, Bangzhao Shu*, Hua Shen*, Jiaxin Pei*, Haotian Zhang*, Miriam Schirmer*, David Jurgens (* = Equal Contribution with Random Order)
[Under Review at NAACL 2025]

Research Experience

Human-Centered NLP

Towards Simulating Culturally-Aligned Mental Health Counseling Sep. 2024 - Present
Advisor: Rada Mihalcea

- Simulate counselor-patient conversations in mental health from Spanish culture using LLMs, based on **visual**, **speech**, and **text** information of the dataset in **[S.3]**.
- Align LLMs with doctor and patient behaviors in Spanish culture by exploring strategies such as prompt-based tuning, fine-tuning, **reinforcement learning** (e.g., DPO), and **multi-agent frameworks**.
- Developed a **human-LLM collaborative evaluation framework** to assess cultural sensitivity, behavioral alignment, and potential counseling quality of simulated personas, minimizing reliance on existing QA benchmarks.

Causal Effect of Intervention Messages on Medical Interns' Well-being [S.4] Feb. 2024 - Present

Advisors: Maggie Makar, Rada Mihalcea

- Collaborate with the Sen Lab at Michigan Psychiatry to explore their 4-year data on intervention messages and 6,000+ medical interns' daily behavioral data, including mood, steps, and sleep.
- Measure and model the heterogeneous treatment effects of different categories of intervention messages on various subgroups of patients (e.g. messages incorporating "mindfulness" are effective in improving female interns' sleep), employing tools such as EconML and scikit-learn.
- Develop a method robust to data perturbations to identify subgroups of patients with significant reactions to treatments.

NLP for Enhanced Behavioral Counseling in Spanish [S.3] Apr. 2024 - Oct. 2024

Advisors: Verónica Pérez-Rosas, Rada Mihalcea

- Introduced a new Spanish counseling dataset, created from 74 public video sources and annotated by experts on counseling strategies. The dataset is used in my ongoing work to simulate culturally aware doctor-patient interaction.
- Analyzed **social-linguistic** and **cultural** differences between the Spanish dataset and a parallel English dataset, such as the word-exchange ratio, language usage (LIWC), and sentiment.

Assessing LLMs' Simulation of Human Responses in Dialogue [S.2] July. 2024 - Sep. 2024

Advisors: David Jurgens

- Evaluated **alignment** of LLM simulations with human interactions on 100,000 English, Chinese, and Russian dialogues from the WildChat dataset, finding a low alignment across all three languages.
- Introduced an **evaluation** framework and associated metrics (multilingual/lexical/style) to assess simulation accuracy.

Inspecting Age Bias in the Value Systems of LLMs [C.1] April. 2024 - Jun. 2024

Advisors: Rada Mihalcea

- Analyzed the alignment of social, economic, and other 11 categories of world values across six age groups in 62 countries on leading LLMs, leveraging data from the World Values Survey.
- Responsible for evaluating the Mistral model and studying the impact of age identity prompts on value discrepancies.
- Findings suggested a general inclination of LLM values towards younger demographics, especially when tested on the US population.

Computational Social Science

Climate Change and Socio-Political Stability in Sub-Saharan Africa [🌐] May. 2024 - Present

Advisor: Verónica Pérez-Rosas, Arun Agrawal

- Collaborate with social scientists to analyze the impact of climate and demographic changes on sociopolitical stability in Africa reflected in scientific literature.
- Develop data infrastructure and models for political risk forecasting, collecting, using web scraping tools such as Selenium and Beautiful Soup, and analyzing over 10,000 relevant scientific articles to explore causal relationships between climate change and political instability.

Uncovering the Impact of George Floyd Incident on Podcast Ecosystem [Report] Aug. 2023 - Jun. 2024

Advisors: David Jurgens, Dallas Card

- Modeled the topical trends (using Mallet), political bias, and conversation dynamics in 600,000+ transcribed podcast episodes published from May to June 2020.
- Investigated the impact of George Floyd's death on podcast discussions across 79 themes, identifying named entities mentioned with George Floyd that reflect themes of social justice and police violence.
- Fine-tuned and calibrated transformers (e.g., RoBERTa, MiniLM) to classify news content within podcast transcripts, with plans to implement retrieval-augmented generation for improved accuracy.

Modeling Conversational Dynamics in Email Exchanges [S.1] Mar. 2023 - Jun. 2024

Advisor: David Jurgens

- Analyzed 11.3M emails from the GMANE corpus, identifying and measuring key **conversational dynamics** and **social-linguistic** factors of email exchanges including intimacy, formality, and cogency.
- Created a dataset of 1,800 emails annotated for intents, expectations, and 14 pragmatic features; benchmarked models including logistic regression and zero-shot LLMs.
- Conducted **causal analysis** revealing that social status, argumentation quality, and social connection influence response rates, with social status being the most significant.

Industry Experience

Boardx.us Jun. 2023 - Aug. 2023

Mentor: Mr. Feng Zhang

- Contributed to designing collaborative/interactive digital whiteboards with a focus on improving user experience and managing company data on Microsoft Azure.
- Assisted in integrating **AI chatbots** to inspire creativity in user design and writing.
- Introduced accessibility features, including color-blind-friendly design, to make the product more inclusive.

Presentations

"NLP for Enhanced Behavioral Counseling in Spanish"

- Michigan AI Symposium, Embodied AI [Poster] Oct. 2024 (Ann Arbor, MI)
- e-HAIL Symposium, Generative AI in Healthcare [Poster] Sep. 2024 (Ann Arbor, MI)

"Can LLMs Simulate Human Subject Studies?"

- NLP Reading Group, Michigan AI Lab [Slides] | [Note] Sep. 2024 (Ann Arbor, MI)

"Podcast Study: Uncovering News in Non-News Space"

- David Jurgens Lab Meeting, UMSI [Slides] Feb. 2024 (Ann Arbor, MI)

"Tutorial: Topic Modeling with Mallet"

- David Jurgens Lab Meeting, UMSI [Slides] Nov. 2023 (Ann Arbor, MI)

Honours and Awards

University Honors, 2023 - 2024 | University of Michigan [🌐] For achieving a GPA above 3.5 in every full-time term.

IEEE-Eta Kappa Nu, 2022 & 2024 | UC San Diego & University of Michigan [🌐] Invited to membership for ranking in the top 25% of the junior class and top 33% of the senior class

Provost Honors, 2021 - 2022 | UC San Diego [🌐] For achieving a GPA above 3.5 in every full-time term.

Teaching and Leadership

NLP Reading Group, University of Michigan *Co-Organizer* [🌐] Sep. 2024 - Present

- Organize and host weekly paper presentations by designing event setups and providing speaker tutorials.
- Communicate NLP research to attendees from non-CS fields (such as Medicine and Psychology) and undergraduates with limited experience, building an inclusive community of 180 active members.

Volunteer Tutor for College Applications Oct. 2021 - Mar. 2023

- Supported about 10 financially challenged international students from countries such as Mexico and Bangladesh through their first-year or transfer college applications.
- Provided constructive feedback on essay brainstorming, standardized test preparation, major and college choice, etc.

Skills

DL/ML Programming	PyTorch, TensorFlow, PyTorch-Transformers, Scikit-Learn
Programming Languages	Python, C/C++, Java, HTML/CSS
Natural Languages	Hunanese (native), Mandarin (native), English (fluent), Cantonese (conversational), German (basic), Spanish (learning), Serbian (learning)