

# Wenbo Zhang

4TH YEAR PHD STUDENT · INFORMATICS

E343, Westgate Building, Pennsylvania State University (University Park), State College, PA, 16802

☎ +1 814-852-9399 | ✉ wenbozhangjs@gmail.com | 🌐 <https://wenbozhangjs.github.io/>

## Research Interests

Natural Language Processing, Code-mixed Text Processing, Model Alignment, Multilingual Language Model, LLM Agents, AI for Social Impact, Speech Processing.

## Education

### Pennsylvania State University (PSU)

Pennsylvania, USA

DOCTOR OF PHILOSOPHY IN INFORMATICS (GPA: 4.0/4.0)

Aug. 2021 - Present

- Advisor: Dr. Amulya Yadav

### University of Southern California (USC)

California, USA

MASTER OF SCIENCE IN ELECTRICAL ENGINEERING

Aug. 2016 - May. 2018

- Advisor: Dr. Cauligi Raghavendra

### University of Electronic Science and Technology of China (UESTC)

Sichuan, China

BACHELOR OF ENGINEERING IN RENEWABLE ENERGY MATERIALS AND DEVICES

Sept. 2011 - Jul. 2015

## Publications

### PEER REVIEWED PUBLICATIONS

- [1] Xinyu Wang, **Wenbo Zhang**, Sai Koneru, Hangzhi Guo, Bonam Mingole, S. Shyam Sundar, Sarah Rajtmajer, Amulya Yadav. The Reopening of Pandora's Box: Analyzing the Role of LLMs in the Evolving Battle Against AI-Generated Fake News. In Proceedings of the 2025 Annual Conference of the North American Chapter of the Association for Computational Linguistics (NAACL), 2025.
- [2] Hangzhi Guo, Xinchang Xiong, **Wenbo Zhang**, Amulya Yadav. ReLax: Efficient and Scalable Recourse Explanation Benchmarking using JAX. In Journal of Open Source Software (JOSS), 2024.
- [3] **Wenbo, Zhang**, Hangzhi Guo, Prerna Ranganathan, Jay Patel, Sathyanath Rajasekharan, Nidhi Danayak, Manan Gupta, Amulya Yadav. A Continual Pre-training Approach to Tele-Triaging Pregnant Women in Kenya. In Proceedings of the 37th AAAI Conference on Artificial Intelligence, 2023. (**System has been full-time deployed by Jacaranda Health**)
- [4] Hangzhi Guo, Xinchang Xiong, **Wenbo Zhang**, Amulya Yadav. Efficient and Scalable Recourse Explanation Benchmark using JAX. In the NeurIPS 2023 Workshop on XAI in Action: Past, Present, and Future Applications, 2023.
- [5] **Wenbo Zhang**, Hangzhi Guo, Prerna Ranganathan, Jay Patel, Sathyanath Rajasekharan, Nidhi Danayak, Manan Gupta, Amulya Yadav. TRIM-AI: Harnessing Language Models for Providing Timely Maternal & Neonatal Care in Low-Resource Countries. In the AAAI Workshop on AI for Social Good (AI4SG), 2023.

### ARCHIVAL

- [1] Xinyu Wang\*, **Wenbo Zhang\***, Sarah Rajtmajer. Monolingual and Multilingual Misinformation Detection for Low-Resource Languages: A Comprehensive Survey. arXiv preprint arXiv:2410.18390.
- [2] Hangzhi Guo, Pranav Narayanan Venkit, Eunhae Jang, Mukund Srinath, **Wenbo Zhang**, Bonam Mingole, Vipul Gupta, Kush R. Varshney, S. Shyam Sundar, Amulya Yadav. Hey GPT, Can You be More Racist? Analysis from Crowdsourced Attempts to Elicit Biased Content from Generative AI. arXiv preprint arXiv:2410.15467.
- [3] **Wenbo Zhang**, Hangzhi Guo, Ian Kivlichan, Vinodkumar Prabhakaran, Davis Yadav and Amulya Yadav. A Taxonomy of Rater Disagreement: Surveying Challenges & Opportunities from the Perspective of Annotating Online Toxicity. arXiv preprint arXiv:2311.04345.

## IN PREP

- [1] **Wenbo Zhang**, Aditya Majumdar, and Amulya Yadav. Code-mixed LLM: Improve Large Language Models' Capability to Handle Code-Mixing through Reinforcement Learning from AI Feedback.
- [2] **Wenbo Zhang**, Amulya Yadav. Benchmarking the Ability of LLMs across Wide Variety of Code-Mixed Tasks

## Research Experience

---

### **Adapt LLMs to the scenarios of code-mixed contents understanding through RLHF**

*Pennsylvania, USA*

ADVISOR: DR. AMULYA YADAV

*Apr. 2024 - Present*

- This work attempts to answer the question: Do large language models understand code-mixed contents well? We focus on analyzing the LLMs' ability to understand and deal with code-mixed text between prompt-based models and small fine-tuned based models specifically designed for code-mixed contents processing. We further enhance the existing LLMs better to understand code-mixed contents through reinforcement learning from AI feedback (RLAIF).

### **Assist case assessment of online sexual abuse and exploitation among children**

*Pennsylvania, USA*

ADVISOR: DR. AMULYA YADAV

*Feb. 2024 - Present*

- Children in the Philippines are vulnerable to online sexual abuse and bullying. This research aims to tackle the following research question in partnership with World Hope International-Philippines (WHI-PH): Can we build an AI-driven decision support system which can automate the preliminary triage and case assessments of online sexual abuse and exploitation of children survivors?

### **Machine learning for phenotypic pattern identification of adolescents with drug usage**

*Pennsylvania, USA*

ADVISOR: DR. AMULYA YADAV

*Feb. 2023 - Jul. 2023*

- This research focuses on identifying potential students (in high school) who may use alcohol, cigarette or marijuana in the future. We analyze potential patterns which may lead such behaviors through the machine learning perspective.

### **TRIM-AI: Harnessing language models for providing timely maternal health care**

*Pennsylvania, USA*

ADVISOR: DR. AMULYA YADAV

*Sept. 2021 - Jun. 2022*

- Maternal health situation are severe in Kenya, Africa. This work focuses on developing an NLP framework, using multi-lingual pretraining and continual pretraining, to predict the user's medical situation (emergency level) based on code-mixed SMS messages.
- This framework has been deployed inside the PROMPTS (digital health system developed by Jacaranda Health). According to the feedback from Jacaranda Health, this framework reduces the monthly system management cost by 22.8% and PROMPTS helpdesk's workload by ~12%.

## Awards, Fellowships, & Grants

---

2023 **AAAI-23 student scholarship**, AAAI Conference on Artificial Intelligence (AAAI)

**PSU Student Travel Award**, College of Information sciences and technology, Pennsylvania State University

2022 **AI Societal Impact Award**, Center for Artificial Intelligence Foundations and Engineered Systems (CAFÉ) at Pennsylvania State University

2014 **3rd Class of National People's Scholarship (top 15%)**, University of Electronic Science and Technology of China

2013 **3rd Class of National People's Scholarship (top 15%)**, University of Electronic Science and Technology of China

2012 **3rd Class of National People's Scholarship (top 15%)**, University of Electronic Science and Technology of China

## Teaching Experience \_\_\_\_\_

Fall 2023 **DS 442 Artificial Intelligence**, Teaching Assistant at Pennsylvania State University  
Spring 2022 **DS 442 Artificial Intelligence**, Teaching Assistant at Pennsylvania State University  
Spring 2024 **DS 442 Artificial Intelligence**, Teaching Assistant at Pennsylvania State University

## Industrial Experience \_\_\_\_\_

### Machine Learning Engineer

Beijing, China

AI LAB, KINGSOFT CO., LTD.

Jan. 2019 - Jul. 2021

- Applied recent advanced NLP techniques to develop information extraction modules inside knowledge graph.
- Employed NLP seq2seq models and speech processing techniques to construct the text-to-speech (TTS) system.

### DIRECTION 1: KNOWLEDGE GRAPH

#### Open domain knowledge graph construction

Beijing, China

PROJECT PARTICIPANT

Jan. 2021 - Jul. 2021

- Designed modules (name entity recognition and relation extraction) for Chinese knowledge graph construction.
- The knowledge graph has been deployed inside the Kingsoft electronic notebook website.

### DIRECTION 2: SPEECH PROCESSING (ESPECIALLY TTS, GENERATIVE AI FOR SPEECH SYNTHESIS)

#### English multi-speaker text-to-speech (TTS) system for novel website

Beijing, China

PROJECT PARTICIPANT

Jul. 2020 - Dec. 2020

- Developed a system which generated speech with someone's tone through few minutes' voice recordings.
- Created a prototype for audiobook reading on English novel translation website to support multiple human voices.

#### End-to-end framework for Chinese polyphone pronunciation prediction

Beijing, China

PROJECT LEADER

Apr. 2020 - Jul. 2020

- Built end-to-end framework for pronunciation prediction of Chinese polyphone with multi-phonemic values.
- Improved the pronunciation correctness of Chinese speech synthesis system.

#### NLP based Chinese text prosody prediction

Beijing, China

PROJECT LEADER

Jan. 2020 - Mar. 2020

- Modeled the prosody (short pronunciation break among Chinese words) prediction as the sequence tagging problem.
- Improved the naturalness and quality of the synthesized speech generated from Chinese speech synthesis system.

#### End-to-end Chinese text-to-speech (TTS) system

Beijing, China

PROJECT LEADER

Jan. 2019 - Apr. 2020

- Implemented whole pipeline of end-to-end Chinese TTS system, including the text processing module (which extracts semantic information from input sentences), acoustic model (which predicts acoustic features based on the semantic information), and the vocoder model (which transforms acoustic features into speech signals).
- Applied on the Kingsoft policy question answer (QA) system.

## Past Internship and Research Visit \_\_\_\_\_

### Data Science Intern

New York, USA

NEW YORK LIFE INVESTMENTS

May. 2024 - Aug. 2024

- Financial machine translation through LLMs and agentic workflow.
- Benchmarked the financial translation performance of LLMs on NYLIM test set through multiple dimensionalities.
- Explored innovative agentic workflow to improve the translation quality and increased 1 BLEU score.

### Research Intern

Suzhou, China

COMPREHEND INFORMATION TECHNOLOGY CO., LTD.

Jun. 2017 - Aug. 2017

- NLP-based data mining on the traffic data (electronic checkpoints data) accessed from Suzhou City Brain.
- Predicted and partitioned whole city into different function regions (education area, central business area etc.).

## Skills

---

<b>DevOps</b>	Google Cloud Platform, Amazon Web Services, Alibaba cloud, Docker
<b>Back-end</b>	Django
<b>Programming</b>	Python, R, C, LaTeX, Shell
<b>Framework</b>	Tensorflow, Pytorch