Yijing Zhang

Research Interests *Data-centric Machine Learning, NLP, Foundation Models, Data Efficiency, Robustness*

Contact Information *E-mail:* zhangyijing2002@gmail.com
Website: https://yijingz02.github.io/

Education

B.S. in Computer Science University of Wisconsin - Madison **Sep 2021 - May 2024**
- with Honor in the Major, Computer Science

Publications

Yijing Zhang, Frederic Sala. *Methods for Domain-specific Fine-tuning for Generative Models.* Senior honor thesis for Honors in the Major, L&S Honors Program at University of Wisconsin - Madison, 2024. [Paper]

Lin Zhang, Shentong Mo, **Yijing Zhang**, Pedro Morgado. *Audio-Guided Visual Animation.* Accepted for oral presentation by European Conference on Computer Vision(ECCV) 2024. [Paper] [Code]

Dyah Adila, Changho Shin, **Yijing Zhang**, Frederic Sala. *Can Language Models Safeguard Themselves, Instantly and For Free?* Accepted by International Conference on Machine Learning(ICML) 2024 Workshop on NextGenAISafety. [Paper]

Dyah Adila, Changho Shin, **Yijing Zhang**, Frederic Sala. *Is Free Self-Alignment Possible?* In submission to International Conference on Learning Representations(ICLR) 2025. [Paper]

Yijing Zhang, Dyah Adila, Changho Shin, Frederic Sala. *Personalize LLM: Fake it and Then Align.* In process. In submission to Annual Conference of the North American Chapter of the Association for Computational Linguistics(NAACL) 2025.

Awards & Honors

Honors in the Major, Computer Science. May, 2024
- Thesis: "Methods for Domain-specific Fine-tuning for Generative Model".
- University of Wisconsin - Madison, College of Letters & Science, L&S Honors Program

ACM ICPC 2021 NA Regional Contest: Team Rank 14. 2021
- Competitive programming since middle school. Multiple regional first prize for OI competitions.

Research Projects

On Generating Better Quality Instruction Tuning Data **Sep 2024 - Present**
@Brown University & University of Wisconsin - Madison
- Related Topic: NLP, Generative models.
- Supervisor: Prof. Stephen Bach & Prof. Fred Sala.
- Focus on generating better quality instruction tuning data from unannotated text.

P-AlignEZ (Personalized AlignEZ) **May 2024 - Present**
@University of Wisconsin - Madison
- Related Topic: NLP, Alignment, Personalization.
- Supervisor: Prof. Fred Sala.
- Utilizing AlignEZ approach for large-scale, low-resource personalization for LLMs.
- Planned submission to ICML 2025.

Audio-guided Animation **Mar 2023 - Present**
@University of Wisconsin - Madison

- Related Topic: Computer vision, Generative models, Audio-to-Video.
- Supervisor: Prof. Pedro Morgado.
- Focused on generating audio-video highly synchronized animation with guidance on audio.
- Accepted as oral presentation by ECCV 2024.

AlignEZ **May 2024 - Sep 2024**
@University of Wisconsin - Madison
- Related Topic: NLP, Alignment.
- Supervisor: Prof. Fred Sala.
- Focused on aligning pretrained language models without additional training.
- In submission to ICLR 2025.

Methods for Domain-specific Fine-tuning for Generative Model **Feb 2022 - May 2024**
@University of Wisconsin - Madison
- Related Topics: NLP, Foundation models, Generative models, Fine-tuning, Data Efficiency
- Supervisor: Prof. Fred Sala.
- Served as independent research study for senior honor thesis.
- Focused on investigating the retrainability of synthetic datasets generated by fine-tuned generative models for domain-specific downstream classification tasks and the fine-tuning efficiency for generating higher-quality synthetic datasets.

Experience

University of Wisconsin-Madison, USA **Sep 2024 - Present**
Research Assistant
- Work as research assistant for Prof. Fred Sala.
- Collaborate with Prof. Stephen Bach from Brown University, we focus on synthetic instruction tuning data generation from unannotated data.

University of Wisconsin-Madison, USA **Jan 2022 - May 2024**
Peer Mentor
- Peer Mentor for the course CS400.
- Responsibilities include: Holding drop-in office hours, and answering online questions.
- Wrote a course reference document aimed at enhancing students' comprehension of course materials. The content includes topics such as Java interface design, generics, iterators, etc.

Technical Skills

- **Machine Learning**: Generative models, NLP, Foundation model, GPTs and Computer vision.
- **Math**: Probability, Statistics, Linear Algebra.
- **Research Tools**: Pytorch, TensorFlow, MATLAB, etc.
- **Research Skills**: Experiment design, Data collection, Data analysis, Essay writing etc.
- **Programming languages**: Python (Max length of code for single past project 1500+ lines), Java (∼400 lines), C (∼800 lines), C++ (∼300 lines).
- **Developer skills**: Web Development, Front-end, and Back-end Development