Zifeng Wang

Home Page | Google Scholar | GitHub | zifengw2@illinois.edu

EDUCATION BACKGROUND

University of Illinois Urbana-Champaign — **Illinois, US**
PhD Candidate, Computer Science, The Grainger College of Engineering — Sept. 2021-Present
Research Interest: AI for Healthcare & Drug Development; Advised by: Prof. Jimeng Sun

Tsinghua University — **Shenzhen, China**
MS, Data Science, Tsinghua-Berkeley Shenzhen Institute (TBSI) — Sept. 2018-Jun. 2021
Thesis: Information Bottleneck for Representation Learning: New Vision
Co-advised by: Prof. Shao-Lun Huang, TBSI and Prof. Khalid M. Mosalam, UC-Berkeley

Tongji University — **Shanghai, China**
B.Eng., Civil Engineering — Sept. 2014-Jun. 2018
Advised by: Prof. Suzhen Li

RESEARCH TOPICS

My research focus is AI for drug development and healthcare, including:
- **Prediction**. Flexible individual/trial outcome prediction with tabular/sequence/vision-language models.
- **Generation**. Synthetic EHR/EDC generation to boost AI for health/trial.
- **Optimization**. In silico optimization of clinical trial protocols through retrieval, prediction, & generation.
- **Explanation**. Provide trustworthy explanations to support decision-making in clinical trial design.

PAPERS

♢ **Preprints & Working Papers:**
- **Z Wang**, C Gao, L Glass and J Sun. Artificial Intelligence for In Silico Clinical Trials: A Review. Under submission.
- **Z Wang**, C Xiao, and J Sun. AutoTrial: Prompting Language Models for Clinical Trial Design. Under submission.
- **Z Wang**, C Xiao, and J Sun. SPOT: Sequential Predictive Modeling of Clinical Trial Outcome with Meta-Learning. Under submission.

♢ **Conferences:**
- T Das*, **Z Wang***, and J Sun. TWIN: Personalized Clinical Trial Digital Twin Generation. **KDD'23**.
- **Z Wang**, Z Wu, D Agarwal and J Sun. MedCLIP: Contrastive Learning from Unpaired Medical Images and Text. **EMNLP'22**.
- **Z Wang** and J Sun. PromptEHR: Conditional Electronic Healthcare Records Generation with Prompt Learning. **EMNLP'22**.
- **Z Wang** and J Sun. Trial2Vec: Zero-Shot Clinical Trial Document Similarity Search using Self-Supervision. **Findings of EMNLP'22**.
- **Z Wang** and J Sun. TransTab: Learning Transferable Tabular Transformers Across Tables. **NeurIPS'22**.
- **Z Wang**, R Wen, X Chen, S-L Huang, N Zhang, and Y Zheng. Finding Influential Instances for Distantly Supervised Relation Extraction. **COLING'22 (Oral)**.
- **Z Wang** and J Sun. SurvTRACE: Transformers for Survival Analysis with Competing Events. **ACM-BCB'22**.
- **Z Wang**, S-L Huang, E. E. Kuruoglu, J Sun, X Chen, and Y Zheng. PAC-Bayes Information Bottleneck. **ICLR'22 (Spotlight, 176/3391)**.
- **Z Wang**, Y Yang, R Wen, X Chen, S-L Huang, and Y Zheng. Lifelong Learning Disease Diagnosis on Clinical Notes. **PAKDD'21 (Best Student Paper, 1/768)**. [video]
- **Z Wang**, R Wen, X Chen, S Cao, S-L Huang, B Qian, and Y Zheng. Online Disease Self-diagnosis with Inductive Heterogeneous Graph Convolutional Networks. **WWW'21**. [video]
- **Z Wang**, X Chen, R Wen, S-L Huang, E. E. Kuruoglu, and Y Zheng. Information Theoretic Counterfactual Learning from Missing-Not-At-Random Feedback. **NeurIPS'20**. [poster]
- **Z Wang**, H Zhu, Z Dong, X He, and S-L Huang. Less Is Better: Unweighted Data Subsampling via Influence Function. **AAAI'20**. [poster]

♢ **Journals:**
- **Z Wang**, Y Zhang, K. M. Mosalam, Y Gao, and S-L Huang. Deep Semantic Segmentation for Visual Understanding on Construction Sites. **Computer-Aided Civil And Infrastructure Engineering**, 2021.
- **Z Wang** and S Li. Data-driven Risk Assessment on Urban Pipeline Network Based on a Cluster Model. **Reliability**

Engineering and System Safety, 2020, 196: 106781.

PROFESSIONAL EXPERIENCE

AWS, Amazon, Applied Scientist Intern — May 2023-Aug 2023
Topics: Deep Learning for Drug Discovery
Supervisor: Zichen Wang

Medidata, Dassault Systèmes, Research Intern — Jan 2023-May 2023
Topics: Synthetic Data Generation for Clinical Trials Supervisor: Mandis Beigi

Amplitude, Research Intern — May 2022-Aug. 2022
Topics: Multi-treatment Causal Inference
Supervisor: Cao Xiao

Jarvis Lab, Tencent, Research Intern — Dec. 2019-Jun. 2021
Topics: Information-theoretic DL; Lifelong Learning Diagnostic AI; Graph DL for Clinical Diagnosis
Supervisor: Yefeng Zheng

Noah's Ark Lab, Research Intern — Apr. 2019-Oct. 2019
Topics: Learning from Noisy Data; Unbiased Recommendation;
Supervisors: Zhenhua Dong, Xiuqiang He

PROFESSIONAL SERVICE

- PC Member/Reviewer for NeurIPS'23, ACL'23, KDD'23, IJCAI'23, NeurIPS'22, EMNLP'22, AAAI'22, IJCAI'22, ICIP'21, ICASSP'21.
- Reviewer for TPAMI.

TEACHING

- TA, CS 598 Deep Learning for Healthcare, Prof. Jimeng Sun — *Spring, 2022*
- TA, Optimization Models and Applications, Prof. Laurent El Ghaoui — *Summer, 2020*
- TA, Bayesian Learning and Data Analysis, Prof. Ercan E. Kuruoglu — *Spring, 2020*
- TA, Learning from Data, Prof. Shao-Lun Huang and Prof. Yang Li — *Fall, 2019*

AWARDS

- Yee Memorial Fellowship — *Oct 2022*
- NeurIPS Scholar Award — *Oct 2022*
- Yunni & Maxine Pao Memorial Fellowship — *Feb 2022*
- Outstanding graduate student of Tsinghua University (2/168) — *June 2021*
- Best Student Research Runner-up of 13rd PhD Student Symposium of Bay Area — *June 2021*
- Best Student Paper Award of PAKDD'21 (1/768) — *May 2021*
- National Graduate Student Scholarship at Tsinghua University (3/229) — *Oct. 2020*
- Best Student Research Runner-up of 1st TBSI Workshop On Data Science — *Dec. 2019*
- Outstanding graduate student (4/40), graduate thesis (3/168) of Tongji University — *Jun. 2018*
- Merit student scholarship of Tongji University — *2015/2016/2017*
- Meritorious winner (1st class prize,$\approx$ 7%) in USA Mathematical Contest in Modeling — *Apr. 2017*

SOFTWARE

- **PyTrial**: A Python Package for Artificial Intelligence in Drug Development. [Doc] [Github]
- **TransTab**: Transferable Transformers for Tabular Learning and Prediction. [Doc] [Github]
- **MedCLIP**: Pretrained Vision-Language Model for Medical Images and Reports. [Github]
- **Trial2Vec**: Pretrained Language Model for Clinical Trial Similarity Search. [Github]
- **PromptEHR**: Synthetic EHR Generation with Prompt Learning. [Github]

INVITED TALK

- Medical Vision-Language Modeling from Unpaired Images and Texts, invited by NCBI-NIH — *Apr. 2023*
- Transfer Learning for Tabular Prediction, invited by AI Time — *Feb., 2023*
- Zero-shot Learning and Transfer Learning on Tabular Data, invited by BAAI — *Oct., 2022*
- Zero-shot Learning and Transfer Learning on Tabular Data, invited by AI Time — *Sept., 2022*
- Understanding Deep Learning via Information in Weights, invited by AI Time — *Apr., 2022*
- PAC-Bayes Information Bottleneck, invited by ReadPaper — *Mar., 2022*

(Updated on May 18, 2023.)