

Zhihan Zhou

Add: 1100 Church Street, Evanston, IL, USA 60201 **Tel:** (1) 224-420-0851

Email: zhihanzhou@u.northwestern.edu

[Personal Website](#) / [LinkedIn](#) / [Google Scholar](#)

EDUCATION

Northwestern University

- ✧ Doctor of Philosophy, Computer Science (GPA: **4.0/4.0**) 09/2020 - Present
- ✧ Master of Science, Computer Science (GPA: **4.0/4.0**) 09/2018 - 06/2020

Zhejiang University

- ✧ Bachelor of Science, Mathematics and Applied Mathematics 09/2014 - 06/2018

SELECTED PUBLICATIONS

- ✧ **DNABERT: Pre-trained Bidirectional Encoder Representation from Transformers for DNA Language Modeling in Genome**
Yanyong Ji, Zhihan Zhou*, Han Liu, Ramana V Davuluri (*: joint first author) ([Over 2 million downloads](#))*
Bioinformatics, 2021
- ✧ **DNABERT-2: Efficient Foundation Model and Benchmark for Multi-Species Genome**
Zhihan Zhou, Yanrong Ji, Weijian Li, Pratik Dutta, Ramana Davuluri, Han Liu ([Over 1 million downloads](#))
ICLR 2024
- ✧ **GenomeOcean: Efficient Foundation Model for Genome Generation**
Zhihan Zhou, Weimin Wu*, Jieke Wu, Lizhen Shi, Zhong Wang, Han Liu (*: joint first author)*
Preprint, 2024
- ✧ **DNABERT-S: Learning Species-Aware DNA Embedding with Genome Foundation Model**
Zhihan Zhou, Weimin Wu*, Harrison Ho, Lizhen Shi, Zhong Wang, Han Liu (*: joint first author)*
Preprint, 2024
- ✧ **Learning Dialogue Representations from Consecutive Utterances**
Zhihan Zhou, Dejiao Zhang, Wei Xiao, Nicholas Dingwall, Xiaofei Ma, Andrew Arnold, Bing Xiang
NAACL 2022

WORK EXPERIENCE

- Research Intern @ NVIDIA, Parabricks** 05/2024 – 09/2024
 - ✧ Build long-context sequence model for simultaneous germline and somatic structural variants calling.
- Research Intern @ Snap, User Modeling and Personalization Team** 06/2023 – 10/2023
 - ✧ Develop foundation models to generate general-purpose user embedding for millions of Snap users. Conduct comprehensive research in data representation, model design, training objective, and evaluation pipelines.
 - ✧ Leads to two academical papers and a system that is being developing and deploying to the product.
- Applied Scientist Intern @ Amazon, AWS AI Labs** 09/2021 – 12/2021
 - ✧ Conduct an independent research project to improve few-shot dialogue understanding by large-scale contrastive conversational model pre-training.
 - ✧ Findings and models are published in NAACL 2022.

AWARDS

- ✧ Terminal-year Ph.D. Fellowship @ Northwestern University

RECENT INVITED TALKS

- ✧ **NVIDIA** (July, 2024): present the DNABERT family to a group of scientist and engineers working on “genome on GPU”.
- ✧ **Harvard Medical School** (March, 2024): present the DNABERT-S project on distinguishing different species from genomic sequences.
- ✧ **Stony Brook University** (March, 2024): presents the DNABERT family and discuss the future of genome foundation models.

ACADEMIC SERVICES

- ✧ Serve as Reviewer for: *NeurIPS, ICML, ICLR, ACL, EMNLP, NAACL, AISTATS*, etc.
- ✧ Serve as Teaching Assistant for:
CS 348: Introduction to Artificial Intelligence; CS 397: Seminar in Statistical Language Modeling; CS 496-6: Advanced Topics in Deep Learning, CS 496-9: Statistical Machine learning, CS 497: Deep Learning for Natural Language Processing, Stats 359: Data Science Project

OTHER PUBLICATIONS

- ✧ **USE: Dynamic User Modeling with Stateful Sequence Models**
*Zhihan Zhou**, *Qixiang Fang**, *Leonardo Neves*, *Francesco Barbieri*, *Yozen Liu*, *Han Liu*, *Maarten W. Bos*, *Ron Dotsch* (*: joint first author)
Preprint, 2024
- ✧ **General-Purpose User Modeling with Behavioral Logs: A Case Study with Snapchat**
*Qixiang Fang**, *Zhihan Zhou**, *Leonardo Neves*, *Francesco Barbieri*, *Yozen Liu*, *Dong Nguyen*, *Daniel L. Oberski*, *Maarten W. Bos*, *Ron Dotsch* (*: joint first author)
SIGIR 2024
- ✧ **Efficient Action Robust Reinforcement Learning with Probabilistic Policy Execution Uncertainty**
Guanlin Liu, *Zhihan Zhou*, *Han Liu*, *Lifeng Lai*
Transaction on Machine Learning Research, 2024
- ✧ **KGML-xDTD: A Knowledge Graph-based Machine Learning Framework for Drug Treatment Prediction and Mechanism Description**
Chunyu Ma, *Zhihan Zhou*, *Han Liu*, *David Koslicki*
GigaScience, 2023
- ✧ **Trade the Event: Corporate Events Detection for News-based Event-driven Trading**
Zhihan Zhou, *Liqian Ma*, *Han Liu*
ACL 2021, Findings
- ✧ **Few-shot Slot Tagging with Collapsed Dependency Transfer and Label-enhanced Task-adaptive Projection Network**
Yutai Hou, *Wanxiang Che*, *Yongkui Lai*, *Zhihan Zhou*, *Yijia Liu*, *Han Liu*, *Ting Liu*
ACL 2020
- ✧ **Joint Speaker Diarization and Recognition Using Convolutional and Recurrent Neural Networks**
Zhihan Zhou, *Yichi Zhang*, *Zhiyao Duan*
ICASSP 2018
- ✧ **ELLPMDA: ensemble learning and link prediction for miRNA-disease association prediction**
*Xing Chen**, *Zhihan Zhou**, *Yan Zhao* (*: joint first author)
RNA Biology, 2018

SKILLS

- ✧ Programming languages: Python, Java, R, HTML, CSS
- ✧ Skills: Deep Learning (Pytorch, HuggingFace, DeepSpeed, etc.), Cloud Computing (Kubernetes, Docker, AWS, GCP, etc.), Bioinformatics (pbsim, minimap, BLAST, etc.)