Zhenqiao Song

CMU – Pittsburgh, Pennsylvania, US
626-528-1560 • zhenqias@andrew.cmu.edu • jocelynsong.github.io

Education

Carnegie Mellon University — **Pittsburgh, US**, *Jan. 2024-Now*
- Ph.D. in Language Technologies Institute. Advisor: Prof. Lei Li

University of California, Santa Barbara — **Santa Barbara, US**, *Sep. 2022-Dec. 2023*
- Ph.D. in Computer Science. Advisor: Prof. Lei Li

Fudan University — **Shanghai, China**, *Sep. 2017-Jun. 2020*
- M.S. in Computer Science. Advisor: Prof. Xiaoqing Zheng
- Shanghai Outstanding Graduate Student

Xiamen University — **Xiamen, China**, *Sep. 2013-Jun. 2017*
- B.S. in Computer Science. Advisor: Prof. Jinsong Su
- Ranked **3**/138 peers

Research Interests

My research focuses on the applications of generative AI in scientific domains like biology and chemistry, with a particular emphasis on **protein design**. I'm also interested in **multilingual NLP**, including multilingual machine translation and multilingual text generation.

Employment

Google Deepmind Student Researcher — **New York City, US**, *May. 2025-Now*
Protein Function Team, working with Ankur Parikh and David Belanger
–Research Project on Enzyme Design

NEC Laboratories America Intern — **Princeton, US**, *Jun. 2024-Sep. 2024*
Machine Learning Department, working with Martin Renqiang Min
–Research Project on Protein Complex Design

Broad Institute of Harvard and MIT Visting Scientist Intern — **Boston, US**, *Jun. 2023-Sep. 2023*
Eric and Wendy Schmidt Center, working with Wengong Jin
–Research Project on Protein Design based on Protein Surface

ByteDance AI Lab FTE — **Shanghai, China**, *Jul. 2020- Aug. 2022*
NLP Researcher advised by Dr. Hao Zhou and Dr. Lei Li, MLNLC team.

–Research Work on Deep Generative Models for Natural Language
- switch-GLAT: Multilingual Parallel Machine Translation Via Code-Switch Decoder
- MTG: A Benchmarking Suite for Multilingual Text Generation

–Industrial Applications on Natural Language Generation
- **Multilingual Advertising Title Generation**
  - Implement a multilingual Ad title generation model incorporating eight languages, including English, Chinese, Japanese, Vietnamese, Russian and so on. It is used in an online advertising system of ByteDance, and **the profit arises from $4,000/day to $200,000/day**.
- **Advertisement Classification System**
  - Implement an advertisement classification model incorporating over 200 categories. **It achieves over 90% online average accuracy on all categories and the profit is over 4 million dollars one year.**

ByteDance AI Lab Intern — **Shanghai, China**, *Jun. 2019-Jun. 2020*

Student Researcher advised by Jiaze Chen

–Research Work on Deep Generative Models for advertising bidding

- Triangular Bidword Generation for Sponsored Search Auction
  - The proposed model is applied in the search system of ByteDance and achieves over $300 million every year.

Publications

[1] **Zhenqiao Song**, Ramith Hettiarachchi, Chuan Li, Jianwen Xie, and Lei Li. "Natural Language Guided Ligand-Binding Protein Design". In: *Preprint*. 2025.

[2] Qihao Duan, Bingding Huang, **Zhenqiao Song**, Irina Lehmann, Lei Gu Gu, Rol Eils, and Benjamin Wild. "JanusDNA: A Powerful Bi-directional Hybrid DNA Foundation Model". In: *Preprint*. 2025.

[3] **Zhenqiao Song**, Tianxiao Li, Lei Li, and Martin Renqiang Min. "PPDiff: Diffusing in Hybrid Sequence-Structure Space for Protein-Protein Complex Design". In: *Forty-second International Conference on Machine Learning* (***ICML***). 2025.

[4] **Zhenqiao Song**, Yunlong Zhao, Wengong Jin, Wenxian Shi, Yang Yang, and Lei Li. "Generative Enzyme Design Guided by Functionally Important Sites and Small-Molecule Substrates". In: *Forty-first International Conference on Machine Learning* (***ICML***). 2024.

[5] **Zhenqiao Song**, Tinglin Huang, Lei Li, and Wengong Jin. "SurfPro: Functional Protein Design Based on Continuous Surface". In: *Forty-first International Conference on Machine Learning* (***ICML***). 2024.

[6] Tinglin Huang, **Zhenqiao Song**, Rex Ying, and Wengong Jin. "Protein-Nucleic Acid Complex Modeling with Frame Averaging Transformer". In: *Thirty-eighth Conference on Neural Information Processing Systems* (***NeurIPS***). 2024.

[7] Kexun Zhang, Yee Man Choi, **Zhenqiao Song**, Taiqi He, William Yang Wang, and Lei Li. "Hire a Linguist!: Learning Endangered Languages with In-Context Linguistic Descriptions". In: *Proceedings of the 62th Annual Meeting of the Association for Computational Linguistics* (***ACL***). 2024.

[8] **Zhenqiao Song** and Lei Li. "Importance Weighted Variational Bayes for Protein Sequence Design". In: *Proceedings of the 2023 International Conference on Machine Learning* (***ICML***)*, 2023*. 2023.

[9] **Zhenqiao Song**, Yunlong Zhao, Wenxian Shi, Yang Yang, and Lei Li. "Functional Geometry Guided Protein Sequence and Backbone Structure Co-Design". In: *arXiv preprint arXiv:2310.04343* (2023).

[10] **Zhenqiao Song**, Yunlong Zhao, Yufei Song, Wenxian Shi, Yang Yang, and Lei Li. "Joint Design of Protein Sequence and Structure based on Motifs". In: *arXiv preprint arXiv:2310.02546* (2023).

[11] Wenda Xu, Danqing Wang, Liangming Pan, **Zhenqiao Song**, Markus Freitag, William Yang Wang, and Lei Li. "Instructscore: Towards explainable text generation evaluation with automatic feedback". In: *Proceedings of the 2023 EMNLP* (***EMNLP***). 2023.

[12] **Zhenqiao Song**, Hao Zhou, Lihua Qian, Jingjing Xu, Shanbo Cheng, Mingxuan Wang, and Lei Li. "switch-GLAT: Multilingual Parallel Machine Translation Via Code-Switch Decoder". In: *International Conference on Learning Representations* (***ICLR***). 2022.

[13] Yiran Chen, **Zhenqiao Song***, Xianze Wu, Danqing Wang, Jingjing Xu, Jiaze Chen, Hao Zhou, and Lei Li. "MTG: A Benchmarking Suite for Multilingual Text Generation". In: *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (***NAACL Findings***). 2022.

[14] **Zhenqiao Song**, Jiaze Chen, Hao Zhou, and Lei Li. "Triangular Bidword Generation for Sponsored Search Auction". In: *Proceedings of the 14th ACM International Conference on Web Search and Data Mining* (***WSDM***). 2021, pp. 707–715.

[15] **Zhenqiao Song**, Xiaoqing Zheng, and Xuanjing Huang. "Jointly learning bilingual word embeddings and alignments". In: *Machine Translation* (2021), pp. 1–19.

[16] **Zhenqiao Song**, Xiaoqing Zheng, Lu Liu, Mu Xu, and Xuan-Jing Huang. "Generating responses with a specific emotion in dialog". In: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics* (***ACL***). 2019, pp. 3685–3695.

[17] Lu Liu, **Zhenqiao Song**, and Xiaoqing Zheng. "Improving Coreference Resolution by Leveraging Entity-Centric Features with Graph Neural Networks and Second-order Inference". In: *arXiv preprint arXiv:2009.04639* (2020).

[18] Jinsong Su, **Zhenqiao Song**, Yaojie Lu, Mu Xu, Changxing Wu, and Yidong Chen. "Exploring Implicit Semantic Constraints for Bilingual Word Embeddings". In: *Neural Processing Letters* 48.2 (2018), pp. 1073–1088.

Honors and Awards

- **China National Scholarship**. 2020
- **Shanghai Outstanding Graduate Student Award**. 2020
- **Fudan University Excellent Scholarship**. 2018 - 2019
- **Xiamen University Excellent Scholarship**. 2014 - 2017

Academic Services

- **Journal Reviewer**: Nature Communication, Nature Machine Intelligence, Medicine, TMLR
- **Conference Reviewer**: ACL 2020-2024, EMNLP 2020-2024, NLPCC 2022-2023, IJCAI 2023-2024, ICML2023-2024, NeurIPS 2023-2024, ICLR 2023-2024, COLM 2024
- Co-organizer of the first Generative AI and Biology workshop (GenBio) in NeurIPS 2023
- Co-organizer of the 2022 SoCal (Southern California) NLP Symposium

Students Advising

- Graduate Students
  - Jacky Chen (2024.9-2025.1, University of Pitt PhD)
  - Ramith Hettiarachchi (2024.9-2025.1, CMU PhD)
  - Xiwei Cheng (2024.6-2025.4, UCSD master -> NEU PhD)
  - Yujia Gao (2024.1-2024.5, CMU master)
- Undergraduate Students
  - Junting Zhou (Peking University, 2025.6 – Now)
  - Charles Novak (CMU, 2025.6 – Now)
  - Jingyu Zhu (Peking University, 2024.6-2024.9)
  - Yufei Song, 2023.1-2023.12, UCSB -> UCLA master)
- Interns at ByteDance AI Lab
  - Lu Liu (Fudan University Master, 2021.6 - 2021.9, now ByteDance NLP researcher)
  - Yiran Chen (Fudan University Master, 2021.6 - 2021.9, now startup NLP researcher)

Invited Talk

- Apr. 2025: Invited talk at Tsinghua University Air Institute GenSI open day!
- Aug. 2024: Invited talk at Tsinghua University Air Institute!

- Jul. 2024: Invited talk at Fudan University NLP Group!
- Oct. 2023: Invited talk at Jiangmen!
- May 2023: Invited talk at BytaDance Research!

Teaching

- **Teaching Assistant** at UCSB: CS 190I Deep Learning at Winter 2023; CS 165B Machine Learning at Spring 2023
- **Teaching Assistant** at Fudan University: Data Structures and Algorithms; Linear Algebra

Skill Set

- **Machine Learning API:** Tensorflow, Pytorch
- **IELTS:** 7.0