

# Danqing Wang

Artificial Intelligence - CMU LTI - PA

✉ dqwang122@gmail.com • 📄 dqwang122.github.io

**Research Interest:** Interested in various generation tasks, including natural language and biological sequence. Devoted to helping the machine generate new content that benefits human life. Mainly focus on (but not limited to):

- Natural Language: Generative Agent, Human Alignment, Controllable Generation, etc.
- Biological Sequence: Peptide Design, Antibody Discovery, etc.

**Github:** <https://github.com/dqwang122>

## Education

- **PhD in Artificial Intelligence** **CMU LTI**  
2024.1 – Current  
*Advisor: Prof. Lei Li*
- **PhD in Computer Science** **UC, Santa Barbara**  
2022.9 – 2023.12  
*Advisor: Prof. Lei Li*
- **Master in Computer Science** **Fudan University**  
2018.9 – 2021.1  
*Advisor: Prof. Xipeng Qiu and Prof. Xuanjing Huang*  
*Ranking: Top 5%*
- **Bachelor in Computer Science and Technology** **Fudan University**  
2014.9 – 2018.6  
*Ranking: Top 10%*

## Experience

- **Research Scientist Intern** **Meta AI (FAIR)**  
2023.6 – 2024.1  
*Advisor: Yuandong Tian*
- **Research Scientist** **ByteDance Research (AI Lab)**  
2020.4 – 2022.8  
*Advisor: Jiaze Chen, Hao Zhou, and Lei Li*

## Main Publication

- **Learning Personalized Alignment for Evaluating Open-ended Text Generation** 📄 *Preprint on arXiv*  
*Danqing Wang, Kevin Yang, Hanlin Zhu, Xiaomeng Yang, Andrew Cohen, Lei Li, Yuandong Tian*
  - Investigate interpretable personalization in open-ended text generation. Propose two personalized evaluation datasets.
  - Develop a personalized story evaluation model PerSE to give detailed review-specific feedback on the story plot. It achieves a higher correlation with humans than GPT-4 by 15.8% on the Kendall correlation.
- **Learning from Mistakes via Cooperative Study Assistant for Large Language Models** 📄 *EMNLP 2023*  
*Danqing Wang, Lei Li*
  - Propose a general framework SALAM to assist LLM in learning from mistakes through interactive cooperation between the main LLM and the study assistant.
  - Develop the model-specific study assistant via the imitation learning of successful experiences, making feedback better aligned with specific LLM behaviors.
- **Instructscore: Towards Explainable Text Generation Evaluation with Automatic Feedback** 📄 *EMNLP 2023*  
*Wenda Xu, Danqing Wang, Liangming Pan, Zhenqiao Song, Markus Freitag, William Yang Wang, Lei Li*
  - Introduce an interpretable metric INSTRUCTSCORE for text generation tasks. It provides comprehensive evaluation by diagnostic reports, culminating in a well-justified final score.
  - It significantly surpasses strong baselines in four generation tasks and achieves the best performance in unseen keyword-to-dialogue generation.
- **ALGO: Synthesizing Algorithmic Programs with Generated Oracle Verifiers** 📄 *NeurIPS 2023*  
*Kexun Zhang, Danqing Wang, Jingtao Xia, William Yang Wang, Lei Li*
  - Introduce a novel framework for algorithm synthesis (ALGO) verified by LLM-generated oracles.
  - Equipped with ALGO, we achieve an 8× better one-submission pass rate over the Codex model and a 2.6× better one-submission pass rate over CodeT.
- **On Pre-training Language Model for Antibody** 📄 *ICLR 2023*  
*Danqing Wang, Fei Ye, Hao Zhou*

- The first comprehensive antibody benchmark: AnTibody Understanding Evaluation (ATUE).
- Explore the representation capability of pre-trained language models on antibody problems with different specificity, especially the influence of the introduction of evolutionary information (EATLM).
- **Accelerating Antimicrobial Peptide Discovery with Latent Structure** 🦋 **KDD 2023**  
*Danqing Wang, Zeyu Wen, Fei Ye, Lei Li, Hao Zhou*
  - Sample peptides from the latent secondary structure space to control the peptide properties.
  - The generated peptides have a high AMP probability (93.62%) and 2/21 show high activity in wet laboratory experiments.
- **Contrastive Aligned Joint Learning for Multilingual Summarization** 🦋 **ACL 2021 Finding**  
*Danqing Wang, Jiaze Chen, Hao Zhou, Xipeng Qiu, Lei Li*
  - A large-scale multilingual summarization corpus MLGSum with 1.1 million articles and summaries in 12 languages.
  - Propose two tasks, contrastive sentence ranking and sentence-aligned substitution, for multilingual summarization.
- **Heterogeneous Graph Neural Networks for Extractive Document Summarization** 🦋 **ACL 2020**  
*Danqing Wang\*, Pengfei Liu\*, Yining Zheng, Xipeng Qiu and Xuanjing Huang*
  - Introduce word nodes to model the cross-sentence relationship for extractive summarization.
  - Easily adapt the graph model from single to multiple document summarization.
- **Extractive Summarization as Text Matching** 🦋 **ACL 2020**  
*Ming Zhong\*, Pengfei Liu\*, Yiran Chen, Danqing Wang, Xipeng Qiu and Xuanjing Huang*
  - Formulate extractive summarization as a semantic text matching problem and select sentences in the summary level.
  - Achieve superior performance on six benchmark datasets, including state-of-the-art extractive results on CNN/DailyMail.
- **Searching for Effective Neural Extractive Summarization: What Works and What's Next** 🦋 **ACL 2019**  
*Ming Zhong\*, Pengfei Liu\*, Danqing Wang, Xipeng Qiu, Xuanjing Huang*
  - Models with an autoregressive decoder are prone to achieving better performance against non-auto-regressive ones.
  - LSTM is more likely to suffer from the architecture overfitting problem while Transformer is more robust.

## Academic Services

---

- Program Committee of ICML, ICLR, NeurIPS, ACL, EMNLP, AAAI
- TA for CS190I Deep Learning (Winter 2023) and CS165B Machine Learning (Spring 2023)
- Local Organization Chair of Social NLP Symposium 2022

## Honor

---

- **Sept 2022** Academic Excellence Fellowship
- **May 2021** Shanghai Outstanding Graduates (5% of graduates)
- **Nov. 2020** Venustech Scholarship (1% of Fudan students)
- **Sept. 2019** Scholarship for Outstanding Students (First Prize)
- **Dec. 2017** Fudan's Undergraduate Research Opportunities Program