Pengyu Cheng

✉ pengyucheng95@gmail.com • 🌐 linear95.github.io

Introduction

I am a researcher at Alibaba Qwen Applications Business Group, leading RL training of the Qwen Large Model Application Team. We primarily focus on enhancing LLMs' foundational capacity via training techniques such as RLHF, RLVR, and agentic RL. I am an enthusiast about LLM self-evolution via multi-agent gaming, which I believe is the most promising path to unleashing the potential of large models. Besides, I have substantial experience in NLP fairness, text generation, representation learning, and information theory.

Education

Duke University 08/2017 – 05/2021
Ph.D., Electrical and Computer Engineering *Adviser: Lawrence Carin*

Tsinghua University 08/2013 – 07/2017
B.S., Mathematics and Statistics *Adviser: Jiwen Lu*

Work Experiences

Alibaba Qwen Applications Business Group 07/2025 – Present
Researcher at Qwen Large Model Application Team *Supervisor: Xiaoxi Jiang*
Leading RL training of foundation models via RLHF, RLVR, and agentic RL.

Moonshot AI 07/2024 – 07/2025
Researcher at RL & Agent Team *Supervisor: Flood Sung, Zhilin Yang*
Training LLM Agents to complete tasks in Web/OS environments via reinforcement learning.

Tencent AI Lab 08/2022 – 07/2024
Senior Researcher at Hunyuan LLM Team *Supervisor: Nan Du, Xiaolong Li, Zhengyou Zhang*
Research and projects on LLM training, AI agents, dialogue systems, and controllable text generation.

Tencent Interactive Entertainment Group 06/2021 – 08/2022
Senior Researcher
Applications of dialogue systems, controllable text generation, style transfer in gaming scenarios and Metaverse.

Internships

Microsoft Cloud & AI 06/2020 – 08/2020
Research Internship *Mentor: Zhe Gan, Yu Cheng, Supervisor: Jingjing Liu*
Improving self-supervised multi-view contrastive learning with learnable data augmentations.

NEC Laboratories America 05/2019 – 08/2019
Research Internship *Mentor: Martin Renqiang Min*
Improving disentangled text representation learning with information-theoretic guidance.

Sogou Map Rendering Group 08/2014 – 09/2014
Research Internship *Mentor: Mao Wang*
Automatic smoothing and compression for polygonal line-like city road data.

Selected Publications

*Equal contribution. †Corresponding Author.

- H. Lu*, Y. Wen*, **P. Cheng**†, et al., *"Search Self-play: Pushing the Frontier of Agent Capability without Supervision"*, ICLR 2026
- Z. Li, **P. Cheng**†, Z. Yu, et al., *"Eliminating Inductive Bias in Reward Models with Information-Theoretic Guidance"*, ICLR 2026
- Y. Du*, Z. Li*, **P. Cheng***, et al., *"Simplify RLHF as Reward-weighted SFT: A Variational Method"*, TMLR 2026
- **P. Cheng**, T. Hu, H. Xu, et al., *"Self-playing Adversarial Language Game Enhances LLM Reasoning"*, NeurIPS 2024
- J. Xie, **P. Cheng**†, et al., *"Chunk, Align, Select: A Simple Long-sequence Processing Method for Transformers"*, ACL 2024

- **P. Cheng***, Y. Yang*, J. Li*, et al., *"Adversarial Preference Optimization: Enhancing Your Alignment via RM-LLM Game"*, ACL 2024
- D. Zeng*, Y. Dai*, **P. Cheng***, et al., *"On Diversified Preferences of Large Language Model Alignment"*, EMNLP 2024
- K. Bai*, **P. Cheng***, W. Hao, et al., *"Estimating Total Correlation with Mutual Information Estimators"*, AISTATS 2023
- R. Wang*, **P. Cheng***, R. Henao, *"Mitigating Gender Bias for Text Generation via MI Minimization"*, AISTATS 2023
- **P. Cheng***, W. Hao*, et al., *"FairFil: Contrastive Neural Debiasing Method for Pretrained Text Encoders"*, ICLR 2021
- S. Yuan*, **P. Cheng***, W. Hao, S. Si, et al., *"Improving Zero-Shot Voice Style Transfer via Disentangled Representation Learning"*, ICLR 2021
- **P. Cheng**, W. Hao, S. Dai, et al., *"CLUB: A Contrastive Log-ratio Upper Bound of Mutual Information"*, ICML 2020
- **P. Cheng**, M. Min, D. Shen, et al., *"Improving Disentangled Text Representation Learning with Information-Theoretic Guidance"*, ACL 2020
- **P. Cheng**, Y. Li, X. Zhang, et al., *"Dynamic Embedding on Textual Networks via a Gaussian Process"*, AAAI 2020 Oral
- **P. Cheng***, D. Shen*, D. Sundararaman, et al., *Learning Compressed Sentence Representations for On-Device Text Processing*, ACL 2019 Oral.

Services & Awards

- Area Chair of ARR 2025 & ARR 2026
- Senior Program Committee of IJCAI 2021
- Reviewer/Program Committee of ICML, NeurIPS, ICLR, AAAI, IJCAI, ACL, EMNLP, NAACL, and ARR.
- Fellowship of Electrical and Computer Engineering at Duke 08/2018
- First in Duke-Tsinghua Machine Learning Summer School (1/112) 08/2017
- Academic Excellence Award of Tsinghua University (top 30%) 10/2014
- Top 5 in the 18-th "Sogou Cup" Artificial Intelligence Programming Contest (5/200) 04/2014
- Silver medal in the 28-th Chinese Mathematical Olympiad (CMO) 01/2013
- First Prize in Chinese National Olympiad in Informatics in Provinces (NOIP) 11/2012

Publications

Zhuo Li, Pengyu Cheng, Zhechao Yu, Feifei Tong, Anningzhe Gao, Tsung-Hui Chang, Xiang Wan, Erchao Zhao, Xiaoxi Jiang, and Guanjun Jiang. Eliminating inductive bias in reward models with information-theoretic guidance. In *International Conference on Learning Representations*, 2026.

Hongliang Lu, Yuhang Wen, Pengyu Cheng, Ruijin Ding, Jiaqi Guo, Haotian Xu, Chutian Wang, Haonan Chen, Xiaoxi Jiang, and Guanjun Jiang. Search self-play: Pushing the frontier of agent capability without supervision. In *International Conference on Learning Representations*, 2026.

Yuhao Du, Zhuo Li, Pengyu Cheng, Zhihong Chen, Yuejiao XIE, Xiang Wan, and Anningzhe Gao. Simplify rlhf as reward-weighted sft: A variational method. *Transactions on Machine Learning Research*, 2026.

Kimi Team. Kimi-vl technical report. *arXiv preprint arXiv:2504.07491*, 2025.

Yuhao Du, Zhuo Li, Pengyu Cheng, Xiang Wan, and Anningzhe Gao. Atoxia: Red-teaming large language models with target toxic answers. In *Findings of the Association for Computational Linguistics: NAACL 2025*, pages 3251–3266, 2025.

Pengyu Cheng, Yifan Yang, Jian Li, Yong Dai, Tianhao Hu, Peixin Cao, Nan Du, and Xiaolong Li. Adversarial preference optimization: Enhancing your alignment via rm-llm game. In *Findings of the Association for Computational Linguistics*, 2024.

Jiawen Xie, Pengyu Cheng, Xiao Liang, Yong Dai, and Nan Du. Chunk, align, select: A simple long-sequence processing method for transformers. In *Proceedings of the 62th Annual Meeting of the Association for Computational Linguistics*, 2024.

Pengyu Cheng, Tianhao Hu, Han Xu, Zhisong Zhang, Yong Dai, Lei Han, nan du, and Xiaolong Li. Self-playing

adversarial language game enhances llm reasoning. In *Advances in Neural Information Processing Systems*, volume 37, pages 126515–126543, 2024.

Pengyu Cheng, Jiawen Xie, Ke Bai, Yong Dai, and Nan Du. Everyone deserves a reward: Learning customized human preferences. *arXiv preprint arXiv:2309.03126*, 2023.

Rui Wang, Pengyu Cheng, and Ricardo Henao. Toward fairness in text generation via mutual information minimization based on importance sampling. In *International Conference on Artificial Intelligence and Statistics*, pages 4473–4485, . PMLR, 2023.

Ke Bai, Pengyu Cheng, Weituo Hao, Ricardo Henao, and Larry Carin. Estimating total correlation with mutual information estimators. In *International Conference on Artificial Intelligence and Statistics*, pages 2147–2164, . PMLR, 2023.

Dun Zeng, Yong Dai, Pengyu Cheng, Tianhao Hu, Wanshun Chen, Nan Du, and Zenglin Xu. On diversified preferences of large language model alignment. *arXiv preprint arXiv:2312.07401*, 2023.

Pengyu Cheng and Ruineng Li. Replacing language model for style transfer. *arXiv preprint arXiv:2211.07343*, 2022.

Shengxuan Luo, Pengyu Cheng, and Sheng Yu. Semi-constraint optimal transport for entity alignment with dangling cases. In *Findings of the Association for Computational Linguistics*, pages 2330–2339, 2022.

Weituo Hao, Nikhil Mehta, Kevin J Liang, Pengyu Cheng, Mostafa El-Khamy, and Lawrence Carin. Waffle: Weight anonymized factorization for federated learning. *IEEE Access*, 10:49207–49218, 2022.

Hao Zhang, Long Tian, Zhengjue Wang, Yishi Xu, Pengyu Cheng, Ke Bai, and Bo Chen. Multiscale visual-attribute co-attention for zero-shot image recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2021.

Pengyu Cheng, Weituo Hao, Siyang Yuan, Shijing Si, and Lawrence Carin. Fairfil: Contrastive neural debiasing method for pretrained text encoders. In *International Conference on Learning Representations*, 2020.

Pengyu Cheng, Yitong Li, Xinyuan Zhang, Liqun Chen, David Carlson, and Lawrence Carin. Dynamic embedding on textual networks via a gaussian process. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 7562–7569, 2020.

Pengyu Cheng, Martin Renqiang Min, Dinghan Shen, Christopher Malon, Yizhe Zhang, Yitong Li, and Lawrence Carin. Improving disentangled text representation learning with information-theoretic guidance. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 7530–7541, 2020.

Siyang Yuan, Pengyu Cheng, Ruiyi Zhang, Weituo Hao, Zhe Gan, and Lawrence Carin. Improving zero-shot voice style transfer via disentangled representation learning. In *International Conference on Learning Representations*, 2020.

Pengyu Cheng, Weituo Hao, Shuyang Dai, Jiachang Liu, Zhe Gan, and Lawrence Carin. Club: A contrastive log-ratio upper bound of mutual information. In *International conference on machine learning*, pages 1779–1788, . PMLR, 2020.

Chang Liu, Jingwei Zhuo, Pengyu Cheng, Ruiyi Zhang, and Jun Zhu. Understanding and accelerating particle-based variational inference. In *International Conference on Machine Learning*, pages 4082–4092, . PMLR, 2019.

Liqun Chen, Guoyin Wang, Chenyang Tao, Dinghan Shen, Pengyu Cheng, Xinyuan Zhang, Wenlin Wang, Yizhe Zhang, and Lawrence Carin. Improving textual network embedding with global attention via optimal transport. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5193–5202, 2019.

Dinghan Shen, Pengyu Cheng, Dhanasekar Sundararaman, Xinyuan Zhang, Qian Yang, Meng Tang, Asli Celikyilmaz, and Lawrence Carin. Learning compressed sentence representations for on-device text processing. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 107–116, 2019.

Pengyu Cheng, Chang Liu, Chunyuan Li, Dinghan Shen, Ricardo Henao, and Lawrence Carin. Straight-through estimator as projected wasserstein gradient flow. In *NeurIPS 2018 Bayesian Deep Learning Workshop*, 2018.