Dongqi Cai (蔡栋琪)

PhD Student (Third Year)

School of Computer Science, Beijing University of Posts and Telecommunications, China

Email: cdq@bupt.edu.cn

Homepage: http://www.caidongqi.com/

Research Interests

Federated Learning, Efficient NLP System, Speech Privacy.

Education

09/2024 – 08/2025 **Visiting PhD in Computer Science and Technology**, Cambridge
Advisor: Nicholas D. Lane

09/2021 – present **PhD in Computer Science and Technology**, BUPT

- Advisor: Shangguang Wang, Mengwei Xu
- Remote Advisor: Felix Xiaozhu Lin (University of Virginia)

09/2017 – 07/2021 **BS in Communication Engineering**, BUPT

- Advisor: Lin Fan

Employment & Experience

07/2021 – 12/2021 **Research Intern**, WeBank

- Mentor: Lixin Fan

Honors & Awards

- Travel Grant, EuroSys/MobiSys/ATC, 2024
- National Scholarship, Ministry of Education, 2023
- Outstanding Graduate Student, BUPT, 2023
- Excellent Ph.D. Students Foundation, BUPT, 2023
- Outstanding Graduate Student, State Key Laboratory of Networking and Switching Technology, 2022/2023
- First-class academic scholarship, BUPT, 2022
- National-Level, Innovation and Entrepreneurship Projects for College Students, 2019

Academic Services

- **TPC Member**

MobiSys'24 AE, MobiCom'24 AE, NCSC-edge'22, TURC-SIGBED-China'23

- **Reviewer**

TMC, TKDE, IoTJ, SAGC'22, ICASSP'24.

- **External Reviewer**

EIS'21, ICWS'23

Conference Publications (* = equal contributions; # = corresponding)

(full list at https://scholar.google.com/citations?user=dIimkboAAAAJ&hl=zh-CN)

[C1] "FwdFL: Efficient Federated Finetuning of Language Models"

Mengwei Xu (My advisor), **Dongqi Cai#**, Yaozong Wu, Xiang Li, Shangguang Wang, in *USENIX Annual Technical Conference (**USENIX ATC, CCF-A**),* 2024.

[C2] "FedRDMA: Communication-Efficient Cross-Silo Federated LLM via Chunked RDMA Transmission"

Zeling Zhang*, **Dongqi Cai***, Yiran Zhang, Mengwei Xu, Shangguang Wang, Ao Zhou, in *Proceedings of the 4rd Workshop on Machine Learning and Systems (**EuroMLSys**), co-located with European Conference on Computer Systems (EuroSys, CCF-A),* 2024.

[C3] "Rethinking Mobile AI Ecosystem in the LLM Era"

Jinliang Yuan*, Chen Yang*, **Dongqi Cai***, Shihe Wang, Xin Yuan, Zeling Zhang, Xiang Li, Dingge Zhang, Hanzi Mei, Xianqing Jia, Shangguang Wang, Mengwei Xu, in *Proc. ACM Int. Conf. Mobile Computing and Networking (**MobiCom, CCF-A**)*, 2024.

[C4] "Federated Few-shot Learning for Mobile NLP"

Dongqi Cai, Shangguang Wang, Yaozong Wu, Felix Xiaozhu Lin, Mengwei Xu, in *Proc. ACM Int. Conf. Mobile Computing and Networking (**MobiCom, CCF-A**)*, 2023.

[C5] "Efficient Federated Learning for Modern NLP"

Dongqi Cai, Yaozong Wu, Shangguang Wang, Felix Xiaozhu Lin, Mengwei Xu, in *Proc. ACM Int. Conf. Mobile Computing and Networking (**MobiCom, CCF-A**)*, 2023.

[C6] "Towards Practical Few-shot Federated NLP"

Dongqi Cai, Yaozong Wu, Haitao Yuan, Shangguang Wang, Felix Xiaozhu Lin, Mengwei Xu, in *Proceedings of the 3rd Workshop on Machine Learning and Systems (**EuroMLSys**), co-located with European Conference on Computer Systems (EuroSys, CCF-A),* 2023.

[C7] "GPT4D: Automatic Cross-Version Linux Driver Upgrade Toolkit"

Borui Yang, Hongyu Li, **Dongqi Cai**, in *the 8th EAI International Conference on Machine Learning and Intelligent Communications (MLICOM)*, 2023.

[C8] "FedAdapter: Efficient Federated Learning for Mobile NLP"

Dongqi Cai, Shangguang Wang, Yaozong Wu, Mengwei Xu, in *Proceedings of the ACM Turing Award Celebration Conference (TURC)*, 2023.

[C9] "Towards ubiquitous learning: A first measurement of on-device training performance"

Dongqi Cai, Qipeng Wang, Yuanqiang Liu, Yunxin Liu, Shangguang Wang, Mengwei Xu, in *Proceedings of the 5th International Workshop on Embedded and Mobile Deep Learning* (**EMDL**), *co-located with ACM International Conference on Mobile Systems, Applications, and Services (MobiSys, CCF-B),* 2021.

[C10] "Mitigating App Collusion using Machine Learning"

Xuefei Duan, Hua Lu, Jinliang Yuan, Qiyang Zhang, **Dongqi Cai**, in *IEEE 7th International*

Conference on Big Data Intelligence and Computing (DataCom), 2021.

Journal Publications

[J1] "Accelerating Vertical Federated Learning"

Dongqi Cai, Tao Fan, Yan Kang, Lixin Fan, Mengwei XU, Shangguang Wang, Qiang Yang, early access in *IEEE Transactions on Big Data (**IEEE TBD**),* 2022.

[J2] "Implementation of an E-payment security evaluation system based on quantum blind computing"

Dongqi Cai, Xi Chen, Yuhong Han, Xin Yi, Jinping Jia, Cong Cao, Ling Fan, in *International Journal of Theoretical Physics (IJTP),* 2020.

Patents

[1] "A Federated Learning Method, System, and Apparatus Based on Forward Gradient"

Mengwei Xu; Yaozong Wu; **Dongqi Cai;** Shangguang Wang

[2] "A Federated Few-Shot Learning Method, System, and Device for Natural Language Models"

Mengwei Xu; **Dongqi Cai**; Ao Zhou; Xiao Ma; Shangguang Wang

[3] "A Federated Learning Method, Device, and System for Pre-trained Models"

Mengwei Xu; **Dongqi Cai**; Ao Zhou; Xiao Ma; Shangguang Wang,

[4] "Vertical Federated Learning Modeling Optimization Method, Device, Medium, and Program"

Dongqi Cai; Lixin Fan; Qiang Yang

Invited Talk

- EMDL'21 (Co-located with MobiSys'21), Towards ubiquitous learning: A first measurement of on-device training performance, Online, 2021/06/25
- EuroMLSys'23 (Co-located with EuroSys'23), Towards Practical Few-shot Federated NLP Rome, Italy, 2023/05/08
- MobiCom'23, Efficient Federated Learning for Modern NLP, Madrid, Spain, 2023/10/05
- MobiCom'23, Federated Few-shot Learning for Mobile NLP, 2023/10/05, Madrid, Spain
- 北邮计算机学院（国家示范性软件学院）"砥砺研思，学术领航"学术论坛，面向大语言模型的高效联邦学习系统，北京，中国，2023/12/26
- EuroMLSys'24 (Co-located with EuroSys'24), FedRDMA: Communication-Efficient Cross-Silo Federated LLM via Chunked RDMA Transmission, Athens, Greece, 2024/04/22
- MoiSys'24 N2Women, Large Language Models on Mobile Devices: Measurements, Analysis, and Insights, Tokyo, Japan, 2024/06/03
- EdgeFM'24 (Co-located with MobiSys'24), Large Language Models on Mobile Devices: Measurements, Analysis, and Insights, Tokyo, Japan, 2024/06/07
- USENIX ATC'24, FwdLLM: Efficient Federated Finetuning of Large Language Models with Perturbed Inferences, SANTA CLARA, CA, USA, 2024/11/11
- CCF Talk "计算机网络与通信顶会细读精讲" 主题讲座， 面向大语言模型的高效联邦学习系统，全球直播，时间待定