

XIAO Yaxin

✉ samin-yaxin.xiao@polyu.edu.hk

☎ +852-6207-4608

🌐 <https://scholar.google.com/citations?user=spRkQ2oAAAAJ&hl=en&oi=ao>

HIGHLIGHTS

HKPFS

NeurIPS

AAAI

TDSC

ICCV

RESEARCH AREA

- Model Extraction
- Membership Inference
- Machine Unlearning
- Watermarking Techniques

ACADEMIC SERVICE

- Invited Individual Reviewer, ICCV, ICLR, WWW, CVPR, ICML, IEEE Transactions on Information Forensics and Security (TIFS)
- Reviewer (Under Supervision of Prof. Haibo Hu), ICDE, VLDB



PROFILE

I am currently a Research Assistant Professor in the Department of Electrical and Electronic Engineering at The Hong Kong Polytechnic University, working with Prof. Haibo Hu in the ASTAPLE Group. My research focuses on the privacy and security of machine learning.



WORK EXPERIENCE

Research Assistant Professor	Jun 2026 - Present
The Hong Kong Polytechnic University	
Postdoctoral Researcher	Oct 2025 - May 2026
The Hong Kong Polytechnic University	



EDUCATION

Doctor of Philosophy	Sep 2020 - Sep 2025
The Hong Kong Polytechnic University	
Master of Engineering	Sep 2017 - Jun 2020
Zhejiang University (ZJU)	
Bachelor of Engineering	Sep 2013 - Jun 2017
Zhejiang University (ZJU)	



AWARDS & HONORS

2020 The Hong Kong PhD Fellowship Scheme (HKPFS)
2019 Second-class Ningbo Hanshang Hydraulic Scholarship
2017 Outstanding Graduates of ZJU
2016 National Scholarship of Chinese Government
2016 First-class Scholarship of Outstanding Students of ZJU
2016 Liebherr Enterprise Scholarship of ZJU
2014 Second-class Scholarship of Outstanding Students of ZJU



SELECTED PUBLICATIONS

- [1] **Yaxin. Xiao**, Qingqing Ye, Haibo Hu, Huadi Zheng, Chengfang Fang, and Jie Shi. "MEXML: Pool-based Active Model Extraction Crossover Membership Inference." 36th Conference on Neural Information Processing Systems (NeurIPS), 2022.
- [2] **Yaxin Xiao**, Qingqing Ye, Li Hu, Huadi Zheng, Haibo Hu, Zi Liang, Yijie Jiao, "Privacy Attacks Exploiting Residuals: How Approximate Machine Unlearning Fails to Protect Privacy" The IEEE/CVF International Conference on Computer Vision (ICCV), 2025.
- [3] **Yaxin Xiao**, Haibo Hu, Qingqing Ye, Li Tang, Zi Liang, Huadi Zheng, "Unlocking High-Fidelity Learning: Towards Neuron-Grained Model Extraction". IEEE Transactions on Dependable and Secure Computing (TDSC), 2025.
- [4] **Yaxin Xiao**, Qingqing Ye, Zi Liang, Haoyang Li, RongHua Li, Huadi Zheng, Haibo Hu, "Class-feature Watermark: A Resilient Black-box Watermark Against Model Extraction Attacks" The Annual AAAI Conference on Artificial Intelligence (AAAI), 2026.



OTHER PUBLICATIONS

- [5] Zi Liang, Haibo Hu, Qingqing Ye, **Yaxin Xiao**, RongHua Li. "Does Low Rank Adaptation Lead to Lower Robustness against Training-Time Attacks?" Forty-second International Conference on Machine Learning (ICML), 2025.
- [6] Haoyang Li, Li Bai, Qingqing Ye, Haibo Hu, **Yaxin Xiao**, Huadi Zheng, Jianliang Xu. "A Sample-Level Evaluation and Generative Framework for Model Inversion Attacks". The Annual AAAI Conference on Artificial Intelligence (AAAI), 2025.
- [7] Zi Liang, Pinghui Wang, Ruofei Zhang, Haibo Hu, Shuo Zhang, Qingqing Ye, Nuo Xu, **Yaxin Xiao**, Chen Zhang, Lizhen Cui. "Exploring Intrinsic Alignments within Text Corpus" The Annual AAAI Conference on Artificial Intelligence (AAAI), 2025.
- [8] Tang, Li, Qingqing Ye, Haibo Hu, Qiao Xue, **Yaxin Xiao**, and Jin Li. "DeepMark: A Scalable and Robust Framework for DeepFake Video Detection." ACM Transactions on Privacy and Security 27, no. 1 (2024): 1-26.
- [9] Rong Du, Qingqing Ye, **Yaxin Xiao**, Liantong Yu, Yue Fu, Haibo Hu. "Dual Utilization of Perturbation for Stream Data Publication under Local Differential Privacy". The IEEE 41st International Conference on Data Engineering (ICDE), 2025.
- [10] Zi Liang, Qingqing Ye, Yanyun Wang, Sen Zhang, **Yaxin Xiao**, RongHua Li, Jianliang Xu, Haibo Hu. "Yes, My LoRD. Guiding Language Model Extraction with Locality Reinforced Distillation." Annual Meeting of the Association for Computational Linguistics (ACL), 2025.



PUBLICATIONS UNDER PEER REVIEW

- [1] **Yaxin Xiao**, Yulin Jin, Li Tang, Haoyang Li, Huadi Zheng, Qingqing Ye, Haibo Hu, "Sample-Grained Approximate Unlearning with Provable Per-Sample Bounds". Forty-third International Conference on Machine Learning (ICML), 2026. (Under peer review)
- [2] Kun Fang, Qinghua Tao, Junxu Liu, **Yaxin Xiao**, Qingqing Ye, Jian Sun, Haibo Hu, "Machine Unlearning in Low-dimensional Feature Subspace". Forty-third International Conference on Machine Learning (ICML), 2026. (Under peer review)
- [3] Xiaoyu Xu, Minxin Du, Kun Fang, Zi Liang, **Yaxin Xiao**, Zhicong Huang, Cheng Hong, Qingqing Ye, Haibo Hu. "FIT: Defying Catastrophic Forgetting in Continual LLM Unlearning". Forty-third International Conference on Machine Learning (ICML), 2026. (Under peer review)
- [4] Yulin Jin, **Yaxin Xiao**, Zi Liang, Tairan Huang, Qingqing Ye, Yingnan Zhao, Haibo Hu. "Towards Provable Unlearning Verification via Unlearning Impact Prediction". ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2026. (Under peer review)
- [5] Yulin Jin, Tairan Huang, **Yaxin Xiao**, Zi Liang, Qingqing Ye, Haibo Hu. "Unveiling Learning-Agnostic Generalization Error and it's Relationship to Data Volume by Knowledge Deletion". ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2026. (Under peer review)
- [6] Zi Liang, Haibo Hu, Qingqing Ye, **Yaxin Xiao**, Haoyang Li. "Why Are My Prompts Leaked? Unraveling Prompt Extraction Threats in Customized Large Language Models". Annual Meeting of the Association for Computational Linguistics (ACL), 2026. (Under peer review)
- [7] Haoyang Li, Ruoxi Sun, Qingqing Ye, Benjamin Zi Hao Zhao, Hongsheng Hu, **Yaxin Xiao**, Zi Liang, Xuyun Zhang, Jason (Minhui) Xue, Haibo Hu. "FeatMark: Feature-level Watermark Protection against Mimicry Attacks with Diffusion Models". CCS, 2026. (Under peer review)



TEACHING EXPERIENCE

Teaching Assistant (Departmental Training), The Hong Kong Polytechnic University

- Training Times: 78-100 contact hours per semester.
- Courses: EIE4432 Web Systems and Technologies, EIE4118 Intrusion Detection and Penetration Testing, EIE4117 Capstone Projects, EIE3117 Integrated Project, EIE4435 Digital Video Production and Broadcasting.

Laboratory Instructor, The Hong Kong Polytechnic University

- EIE4432 Web Systems and Technologies (total 266 lab contact hours): lab supervision, code/demo review, Q&A, and assessment.