Kai Yao

PhD Candidate

Informatics Forum, 10 Crichton Street, EH8 9AB, Edinburgh, UK

kai.yao@ed.ac.uk | kaikaiyao.github.io

Education

- **University of Edinburgh** — Edinburgh, UK
  PhD in Cyber Security, Privacy and Trust — 2023 – Present
  Research Focus: Secure and Trustworthy Machine Learning
  Advisor: Dr. Marc Juarez
  Expected Graduation: Nov 2026
- **Johns Hopkins University** — Baltimore, MD, USA
  MS in Mechanical Engineering — 2020
- **Fudan University** — Shanghai, China
  BS in Theoretical Mechanics — 2017

Publications

- **Yao K**, Juarez M. *Smudged Fingerprints: A Systematic Evaluation of the Robustness of AI Image Fingerprints* To appear in the 4th IEEE Secure and Trustworthy Machine Learning Conference, 2026 (Accepted).
- **Yao K**, Juarez M. *AuthPrint: Fingerprinting Generative Models Against Malicious Model Providers.* arXiv preprint, 2025.
- **Yao K**, Juarez M. *SoK: What Makes Private Learning Unfair?* Proceedings of the 3rd IEEE Secure and Trustworthy Machine Learning Conference, 2025.
- Rochman ND*, **Yao K***, Gonzalez NA*, Wirtz D, Sun SX. *Single-Cell Volume Measurement Utilizing the Fluorescence Exclusion Method (FXm).* Bio-protocol. 2020 Jun 20;10(12):e3652.
- **Yao K***, Rochman ND*, Sun SX. *CTRL: A Label-Free Artificial Intelligence Method for Dynamic Measurement of Single-Cell Volume.* Journal of Cell Science. 2020 Apr 1;133(7):jcs245050.
- Perez-Gonzalez NA*, Rochman ND*, **Yao K***, Tao J, Le MT, Flanary S, Sablich L, Toler B, Crentsil E, Takaesu F, Lambrus B. *YAP and TAZ Regulate Cell Volume.* Journal of Cell Biology. 2019 Oct 7;218(10):3472–88.
- **Yao K***, Rochman ND*, Sun SX. *Cell Type Classification and Unsupervised Morphological Phenotyping from Low-Resolution Images Using Deep Learning.* Scientific Reports. 2019 Sep 17;9(1):1–3.
- Zhang Q, Meng Z, Zhang Y, **Yao K**, Liu J, Zhang Y, Jing L, Yang X, Paliwal N, Meng H, Wang S. *Phantom-Based Experimental Validation of Fast Virtual Deployment of Self-Expandable Stents for Cerebral Aneurysms.* BioMedical Engineering OnLine. 2016 Dec;15(2):431–7.

*Note: * denotes equal contributions (i.e. co-first author).*

Research Experience

- **Smudged Fingerprints: Robustness of AI Image Fingerprinting** University of Edinburgh 2024 – 2025
  - Conducted the first systematic security evaluation of model fingerprint detection for generative image attribution, formalizing white-box and black-box threat models and implementing five removal and forgery attack strategies to benchmark 14 fingerprinting methods across RGB, frequency, and learned-feature domains on 12 state-of-the-art GAN, VAE, and diffusion models; revealed severe robustness gaps, utility–robustness trade-offs, and fundamental limitations of passive fingerprinting for adversarial provenance.
- **Fingerprinting Generative Models Against Malicious Providers** University of Edinburgh 2024 – 2025
  - Pioneered the first adversarial extension of generative model fingerprinting to threat models with malicious providers, developing robust fingerprint extraction methods for GANs and diffusion models that achieve near-zero FPR at 95% TPR under architectural modifications and adversarial attacks; built a comprehensive evaluation framework across models and attack scenarios and released an open-source implementation advancing AI security and model provenance research.

- **Differential Privacy's Disparate Impact in Machine Learning** University of Edinburgh 2023 – 2024
  - Conducted a systematic fairness analysis of differentially private machine learning, developing a taxonomy of disparity-inducing factors across DP mechanisms, model architectures, and data distributions; performed causal analysis identifying dataset size and group distance to decision boundaries as key drivers of DP-induced unfairness, evaluated mitigation strategies, and identified critical open research gaps.
- **3D Single-Cell Morphology and Size Prediction with DL** Johns Hopkins University 2019 – 2020
  - Developed a high-throughput, label-free deep learning pipeline to predict 3D single-cell morphology and size from DIC microscopy images by designing a microfluidic fluorescence exclusion measurement system, implementing image preprocessing algorithms, and training an optimized U-Net-based CNN.
- **Cell Type Classification and Morphological Phenotyping with DL** Johns Hopkins University 2018 – 2019
  - Designed a label-free, high-throughput deep learning framework for cell type classification and morphological phenotyping from low-resolution microscopy images, including automated preprocessing, CNN-based morphological clustering, and analysis of relationships among cell type, density, and cancer cell morphology.

Industrial Experience

- **Senior AI Algorithm Engineer** Intel Corp., Shanghai, China 2021 – 2023
  - Led end-to-end development and optimization of Intel AI software stack, including Neural Coder and Intel Extension for PyTorch, to improve training and inference throughput for PyTorch and TensorFlow workloads on Intel hardware. Developed PyTorch adapter algorithms for Intel Neural Compressor to enable INT8 quantization and optimized inference performance for large-scale AIGC models (e.g., Stable Diffusion). Designed and maintained a comprehensive benchmarking framework to evaluate AI workload performance across Intel and competitor platforms, and collaborated with industry partners (e.g., Alibaba, AWS) to integrate Intel AI solutions into production ecosystems.
- **AI Algorithm Engineer** Huawei Technologies Co., Ltd., Shanghai, China 2020 – 2021
  - Developed ML algorithms for 5G systems, focusing on MU-MIMO features using CNN and RNN. Performed model arch optimization and weight compression to improve training and inference efficiency, and collaborated closely with deployment and validation teams to ensure successful integration and real-world performance.

Awards, Fellowships, & Grants

- Conference Attendance Grant, IEEE SaTML Conference, IEEE, 2024 & 2025
- School of Informatics PhD Full Scholarship, University of Edinburgh, 2023
- Division Achievement Award, Fast Stable Diffusion on Intel CPU, Intel AIA, 2023
- Division Recognition Award, Neural Coder Partnering Alibaba Cloud, Intel CESG SW AI, 2023
- Division Achievement Award, Innovation of Neural Coder, Intel AIA, 2022
- Departmental Research Fellowship, Johns Hopkins University, 2017
- Outstanding Graduate of the Year 2017, Fudan University, 2017
- JASSO Full Scholarship for Exchange Students, Japanese Government, 2014

Teaching Experience

- **Privacy and Security with Machine Learning** University of Edinburgh, Edinburgh, UK 2024 – 2025
  Teaching Assistant and Lab Demonstrator (Dr. Marc Juarez)
- **Privacy and Security with Machine Learning** University of Edinburgh, Edinburgh, UK 2023 – 2024
  Teaching Assistant and Lab Demonstrator (Dr. Marc Juarez)
- **Mathematical Image Analysis** Johns Hopkins University, Baltimore, MD, USA 2019 – 2020
  Teaching Assistant (Dr. Mario Micheli)

Student Tutoring Experience

Provided research mentoring or tutoring for: Felipe Takaesu, Eliana Crentsil, Lucia Sablich (Johns Hopkins UG); Shannon Flanary (Johns Hopkins PG); Chunhan Fang (University of Edinburgh PG).