

Research Interest

My research focuses on **AI security**, particularly **AI agent RAG systems**, and **data governance**. I investigate adversarial attacks, practical defenses for secure AI systems, and audio watermarking. I **spent 5 years in industry as a data scientist and product manager**, working on semiconductor yield analytics and forecasting systems.

Education

- Ph.D. in Computer Science, **Indiana University Bloomington**, Bloomington, USA
Aug 2026 - Present
- Ph.D. in Electrical and Computer Engineering, **University of Hawaii at Manoa**, Honolulu, USA
Aug 2024 - Aug 2026 | GPA: 4.00/4.00
- M.S. in Business Analytics, **University of Miami**, Coral Gables, USA
Jul 2018 - Jun 2019 | GPA: 3.64/4.00
- B.A. (Hons) in Accounting and Finance, University of the West of England, Bristol, UK
Aug 2016 - Jun 2018 | GPA: 4.00/4.00
- B.E. in Finance, Guangdong University of Finance, Guangzhou, China
Sep 2014 - Jun 2018 | GPA: 3.40/4.00

Professional Experience

Research Intern | University of Southern California – ISI, Marina Del Rey, USA

Jun 2026 – Aug 2026

- Reproduced and ported 34 research artifacts across 9 topics from ACSAC, IEEE S&P, USENIX, and NDSS to the SPHERE testbed, enabling researchers to reuse the artifacts.
- Built an agent-based tool to screen research artifacts for incompatible hardware and software requirements automatically.

Data Scientist | Joint Laboratory of Materials Science, Xi'an Jiaotong University - Shanghai

Hongzhiwei | Xi'an, China

Jan 2022 – Jul 2024

- Developed a machine learning chip-test classification algorithm with 98.5% accuracy, reducing workload by 75%.
- Investigated process anomalies and yield-loss root causes by correlating production and test data using t-tests, ANOVA, and XGBoost feature importance, contributing to a 10% production-yield improvement within one month.
- Analyzed wafer-map data using DBSCAN to localize spatial defect patterns and developed a CNN-based wafer-map classifier for automated defect identification.
- Designed the architecture and interactive interface for a yield-analysis platform supporting production-data analysis, visualization, and yield investigation.

Product Manager | YMTC FDC Team - Hongzhiwei | Wuhan, China

Jan 2023 – Jul 2024

- Designed the architecture and interactive interface for a Yield Management System supporting production data analysis, visualization, and yield investigation.

- Built a Defect Management System (DMS) and a Wafer Classification System.
- Identified manufacturing and yield trends by analyzing wafer-defect maps and WIP/CP/FT data using the visualization modules.
- Built four statistical process monitoring algorithms based on WECO rules, including consecutive N-point threshold and centerline-side rules, to detect process anomalies.

Data Scientist | Perry Ellis International | Doral, USA

Jun 2019 – Jun 2020

- Built ensemble sales forecasting models with Seasonal ARIMA and LSTM models on Google Cloud Platform, reducing inventory costs by 12%.
- Developed Bayesian sentiment analysis models using word-count and TF-IDF features to identify drivers of customer feedback.
- Built ETL and automated ML prediction pipelines using Apache Beam and Google Cloud Dataflow, supporting scalable batch data processing and model deployment.

Publications

- **[EdgeSP'26]** EchoGuard: Continuous Ultrasound-based Human Presence Verification Against AI Agents
Yizhu Wen, Dharani Nadendla, Shirui Cao, Renzhi Hao, Dong Li, Hanqing Guo
- **[EdgeSP'26]** ULTRASOUND WATERMARK: Real-time Acoustic Watermarking against Voice Scams on Smartphones
Renzhi Hao, Shirui Cao, Yizhu Wen, Rishav Gupta, Dharani Nadendla, Mehran Kafai, Hanqing Guo, Dong Li
- **[ICML'26]** Position: Generative Engine Optimization Creates Underexamined Risks, Governance Must Target Concentration, Disclosure, and Academic Blind Spots [[PDF](#)]
Yizhu Wen, Nan Zhang, Haohan Yuan, Xun Chen, Haopeng Zhang, Hanqing Guo
- **[IEEE S&P'26 Workshops]** Mental Damage: Caption Poisoning Attacks on Retrieval-Augmented Text-to-Music Generation [[PDF](#)]
Yizhu Wen, Shuhao Zhang, Nan Zhang, Long Cheng, Hanqing Guo
- **[EMNLP'25]** Web Intellectual Property at Risk: Preventing Unauthorized Real-Time Retrieval by Large Language Models [[PDF](#)]
Yisheng Zhong, Yizhu Wen, Junfeng Guo, Mehran Kafai, Heng Huang, Hanqing Guo, Zhuangdi Zhu
- **[ICNC'25]** Secure-IRS: Defending Against Adversarial Physical-Layer Sensing in ISAC System [[PDF](#)] [[Demo](#)] [[Slide](#)]
Ziyu Chen, Denny Landika, Alvin Yang, Yizhu Wen, Haofan Cai, Yao Zheng, Hanqing Guo
- **[ICME'24]** DiffImpute: Tabular Data Imputation with Denoising Diffusion Probabilistic Model [[PDF](#)]
Yizhu Wen, Kai Yi, Jing Ke, Yiqin Shen
- **[IEEE EDL'24]** Independent Modulation of Discrete-Time Markov Chain Signals Generated from Tunneling Magnetoresistance Effect [[PDF](#)]

Xihui Yuan, Jiajia Jian, Zheng Chai, Hengjia Wei, Yizhu Wen, Yadong Liu *et al.*

- **[Under review NDSS 2027]** SoK: Is Audio Watermarking Robust to Removal Attacks? [*First author*]
- **[Under review ICASSP 2027]** Real-Time Speech Watermarking for Live Voice Communication: Synthetic Speech Provenance and Post-Hoc Recording Attribution [*First author*]
- **[Under review AAAI 2027]** AuscultRAG: Retrieval-Augmented Audio-LLM for Cardiopulmonary Condition Classification [*Co-first author*]
- **[Under review AAAI 2027]** WAV2PWN: Leveraging Adversarial Transfer Patterns for Surrogate Selection in Black-Box SSL-Based ASR Attacks [*Co-first author*]
- **[Under review ICLR 2027]** PoisonedEar: Knowledge Poisoning Attacks on RAG Multimodal Reasoning in Audio-centric Language Model [*Co-first author*]
- **[Under review ICLR 2027]** Claim–Evidence Prompt Injection for Targeted Answer Contamination in Retrieval-Augmented Generation [*Co-first author*]
- **[Under review ICLR 2027]** Mitigating Cross-Tier Poisoning in Hierarchical Permission-Aware RAG [*Co-first author*]

Technical Skills

Programming Language: Python, R

Visualization: Tableau, Power BI, Spotfire

Cloud Service: AWS, GCP, SPHERE testbed