

Education

- Ph.D. in Electrical and Computer Engineering | **University of Hawaii at Manoa** | Honolulu, USA
Aug 2024 - Present | GPA: 4.00/4.00
Relevant Courses: Large Language Models, Computer Architecture, Digital Systems & Computer Design.
- M.S. in Business Analytics | **University of Miami** | Miami, USA
Jul 2018 - Sep 2019 | GPA: 3.64/4.00
Relevant Courses: Programming for Data Analytics, Data Mining, Machine Learning, Applied Time Series Analysis.
- B.A (Hons) in Accounting and Finance | **University of the West of England** | Bristol, UK
Aug 2016 - Jun 2018 | GPA: 3.71/4.00
- Certificate in C++ In Financial Engineering | **Baruch College, City University of New York** | New York, USA
May 2017 - Aug 2017
Key Topics: Object-Oriented Programming, Standard Template Library, Boost C++ Libraries.
- B.E. in Finance | **Guangdong University of Finance** | Guangzhou, China
Sep 2014 - Jun 2018 | GPA: 3.40/4.00

Publications

- **[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 V 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.*

Posters

- **[USENIX'25]** Real-time Speech Watermark for Defending Hidden Phone Call Recording. [[Poster](#)]
Yizhu Wen, Rui Duan, Yisheng Zhong, Zhuangdi Zhu, Hanqing Guo
- **[USENIX'25]** Web Intellectual Property at Risk: Preventing Unauthorized Real-Time Retrieval by Large Language Models. [[Poster](#)]
Yisheng Zhong, **Yizhu Wen**, Junfeng Guo, Mehran Kafai, Heng Huang, Hanqing Guo, Zhuangdi Zhu
- **[NVIDIA GTC'24]** DiffImpute: Tabular Data Imputation with Denoising Diffusion Probabilistic Model. [[Poster](#)]
Yizhu Wen, Kai Yi, Jing Ke, Yiqin Shen

Research Experience

- Research Assistant (Remote) | **ARCADE Lab, Johns Hopkins University**
May 2023 – Dec 2023

- Research Assistant | **Center for Spintronics and Quantum Systems, Xi'an Jiaotong University**
Aug 2022 – Jul 2024

Professional Experience

- Data Scientist | **Joint Laboratory of Materials Science, Xi'an Jiaotong University - Shanghai Hongzhiwei** | Xi'an, China
Jan 2022 – Jul 2024
 - Engineered a **classification algorithm** for chip test data, achieving a **98.5% accuracy rate** while reducing workload by **75%**.
 - Improved production yield by **10%** within one month through root cause analysis, utilizing statistical models such as **T-tests, ANOVA, and XGBoost feature importance**.
 - Analyzed production parameter Wafer Maps using **DBSCAN clustering** to uncover spatial patterns of abnormal products.
 - Developed a **CNN-based image classification** system for Wafer Map images, enhancing defect detection accuracy.
 - Designed the **system architecture** and interactive interface of the yield analysis platform.
- Product Manager | **YMTC FDC Team - Shanghai Hongzhiwei** | Wuhan, China
Jan 2023 – Jul 2024
 - Completed the **Yield Management System architecture design** for a fab wafer factory, integrating the Defect Management System (DMS) and Wafer Classification and Recognition System (ADC).
 - Collaborated with teams to develop **Defect Wafer Map, Trend Chart, and Map Gallery modules**, as well as data analysis and visualization for WIP/CP/FT.
 - Designed four algorithms based on **WECO rules**, including consecutive N points exceeding the specified standard and consecutive N points on the same side of the centerline.
- Data Scientist | **Perry Ellis International** | Doral, USA
June 2019 – Jun 2020
 - Designed and implemented ensemble models combining **Seasonal ARIMA** and **LSTM** on Google Cloud Platform, achieving a **12% reduction** in inventory costs through accurate sales forecasting.
 - Developed **Bayesian sentiment analysis** models using word count and tf-idf with NLTK, identifying key features influencing customer feedback.
 - Implemented efficient **ETL pipelines** with **Apache Beam** in **Google Cloud Dataflow**, extracting and transforming variables related to time and external factors to enhance data quality for modeling.
 - Deployed machine learning models to **Google Cloud Endpoint**, enabling automated and scalable online batch predictions.
- **Kaggle NYC Taxi Fare Prediction Competition** | Ranking: 32/1488 (top 3%)
Jul 2018 – Sep 2018
 - **Removed outliers** by applying domain knowledge, including fare rate regulations and taxi passenger capacity limits.
 - **Engineered features** such as drop-off distance to landmarks (e.g., Statue of Liberty) and major holiday indicators to enhance model relevance and accuracy.
 - Designed and optimized **model ensembles** and stacking techniques (**XGBoost, LightGBM, Lasso**), effectively minimizing overfitting and improving prediction performance.

Technical Skills

Programming: Python, R, Matlab, C++

Data Science Tools: Scikit-learn, PyTorch, TensorFlow

Cloud Platforms: Google Cloud, AWS

Visualization Tools: Tableau, Spotfire