Hao-Chun (Oscar) Shih

+1(734)-780-4256 / haochuns@umich.edu / GitHub / LinkedIn / Portfolio

ACADEMICS

University of Michigan, Ann Arbor **Aug 2024 – May 2026**
Master of Data Science **GPA: 4.00**
Highlight Coursework: Data Mining, Information Visualization, Machine Learning, Statistical Modeling on Financial Data

National Chung Hsing University **Sep 2019 – Jun 2023**
Bachelor of Science in Applied Mathematics **GPA: 3.92 Ranked 1st**
Highlight Coursework: Mathematical Statistics, Discrete Mathematics, Regression Analysis, Data Structures & Algorithm

WORK EXPERIENCE

Global Quality Planner Intern | (Python, Tableau, MobaXterm, MySQL) **Jul 2025 – Sep 2025**
Micron, GQ Operations Team
- Analyzed over 500,000 historical records of capacity plans and quality testing data to generate robust demand forecasts.
- Enhanced forecast accuracy by 43%, optimizing quality testing plans and resource allocation across operations.
- Developed automation script and created a new database to streamline ETL processes, ensuring end-to-end data retention.
- Built new Tableau dashboards with redesigned UI, enabling real-time monitoring and boosting adoption across 20+ teams.

Research Assistant | (Linux, Python, Latex) **Oct 2023 – Jun 2024**
Academia Sinica, Chinese Knowledge and Information Processing (CKIP) Lab
- Investigated memory capabilities on Llama2-7b with plug-in module Parameter Efficient Fine-Tuning (PEFT)
- Conducted 60+ experiments on number of parameters, pretraining checkpoints, and data paraphrasing effects
- Enhanced retrieval performance by 10% based on F1-Score, through the integration of PEFT and GAR (Generation-Augmented Retrieval), applied to optimize service systems at E-Sun Bank, boosting enterprise productivity by 4%

BI Intern | (SQL, PL/SQL, MySQL, SSIS, DataStage, Cognos, Tableau, PowerBI) **Jul 2022 – Sep 2022**
WPG Holdings, Department of Business Intelligence
- Streamlined cross-departmental workflows by enhancing 27 electronic forms using IBM Cognos, resulting in a 12% increase in processing speed and reducing manual input errors by 8%
- Migrated 54 DataStage databases to SSIS, by implementing optimized ETL processes for industrial decisions, which reduced data processing time by 13% and improved system stability
- Delivered operational improvements that boosted 15% in data accuracy and an estimated annual savings of $50,000 in maintenance and error correction costs

DATA SCIENCE PROJECTS

Transformer-Based Model for Video Segmentation in GUI domain | (Python) [GitHub]
- Finetuned transformer-based temporal models ASFormer and multimodal vision models across over 2,000 videos.
- Engineered a Swin Transformer with spatial attention to enhance feature extraction for high-resolution GUI videos.
- Achieved a 38% increase in action classification accuracy in GUI tasks

Burn-in Board Dashboard Development | (MySQL, Microsoft SQL Server, Tableau) [GitHub]
- Validated data across Tsums, MySQL, and Tableau, resolving discrepancies and improving reporting reliability.
- Redesigned Tableau UI and deployed on a central webpage, improving cross-department access and pattern recognition.
- Rebuilt data pipeline in Python/MySQL, expanding reporting coverage and improving machine allocation accuracy by 16%

Reinforcement Learning for Health Recommendation System | (Python, SQL) [GitHub]
- Developed a personalized health activity recommendation system using Reinforcement Learning, optimizing user activity choices based on past engagement and health metrics
- Balanced exploration and exploitation with Thompson Sampling, adjusting actions to maximize engagement and outcomes.
- Enhanced recommendation accuracy by 25% through integrating contextual factors like user preferences and health data

CNN for Image Classification & RNN for Image Captioning | (Python, NumPy) [Github]
- Engineered a deep learning pipeline from scratch, combining CNN for image extraction and RNN for caption generation
- Achieved 82% image classification accuracy and BLEU-1 score of 0.52 by using mini-batch SGD and Adam optimizer
- Enhanced model stability with batch normalization and dropout; validated correctness via gradient checking

Severe Weather Data Analysis & Visualization Platform | (Altair, Tableau, Streamlit) [GitHub]

- Developed an interactive dashboard for 60,000+ tornado events with dynamic filtering by intensity, damage, and casualties
- Streamlined user experience through Streamlit interface with custom sidebar filters, real-time chart updates, and intuitive layout for seamless stakeholder exploration
- Empowered stakeholders with insights into high-risk states, supporting emergency planning and public resource allocation

Copula-Based Portfolio Risk Modeling & Simulation | (R) [GitHub]

- Built quantitative models to estimate joint tail risk across asset classes using t-Copula, Gumbel, and Clayton copulas, improving dependency modeling accuracy by 30%
- Conducted 10,000+ Monte Carlo simulations to derive portfolio-level Value-at-Risk (VaR) and Expected Shortfall under stress scenarios, enhancing tail risk visibility for high-volatility portfolios
- Calibrated copulas with empirical marginal, boosting risk aggregation robustness and aligning drawdown predictions by 25%

Decisive Factors and Prediction of Health Insurance Premium | (R, Python) [GitHub]

- Improved insurance premium prediction, achieving R^2 of 0.84 through cross-validation
- Enhanced precision using ANOVA, VIF analysis, variable selection, and Box-Cox to address multicollinearity
- Developed Random Forest models, reducing RMSE by 70% in test data, enabling better cost estimations for companies

PUBLICATIONS

Optimizing Knowledge Updates with PEFT in Neural Language Models [pdf]
Hao-Chun Shih, Wei-Yun Ma 2024

- Surveyed Parameter Efficient architectures, investigating the features of FFN-Integrated and Attention-Integrated module
- Revised evaluation methods on memorization ability of model, and quantifying knowledge in LLMs
- Showcased research results at the internationally recognized International Conference on EMNLP

COMPUTER SKILLS

Programming Languages: Python, C++, Java, JavaScript, R, SQL, PL/SQL, HTML, CSS
Database Management: MySQL, Oracle, MongoDB, IBM DataStage, Microsoft SSIS, Cognos, AWS, Google BigQuery
Tools: TensorFlow, Scikit-learn, PyTorch, IBM Cognos, DataStage, SSIS, Jupyter, Latex, Linux, Tableau, Matplotlib, MobaXterm, Django