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 large scale manufacturing data including capacity plans, quality test results, and equipment utilization metrics to improve testing demand forecasts, increasing forecast accuracy by 43% and supporting more efficient resource planning
- Built batch SQL aggregations and automated Python ETL pipelines to transform raw operational data into curated datasets for recurring analysis and reporting
- Delivered Tableau dashboards with operational KPIs used by quality operations and planning teams for real time monitoring, improving alignment on test capacity allocation across 20+ internal teams, including QA, RD, Engineer teams .

Research Assistant | (Linux, Python, Latex) **Oct 2023 – Jun 2024**
Academia Sinica, Chinese Knowledge and Information Processing (CKIP) Lab
- Analyzed performance and cost trade-offs in an enterprise internal knowledge retrieval system by evaluating domain-specific LLMs customization strategies with Parameter Efficient Fine-Tuning (PEFT) technique
- Designed and executed 60+ controlled experiments to assess the impact of model configuration, training checkpoints, and data variations on retrieval quality and response consistency
- Delivered an PEFT and RAG-integrated prototype with clear recommendations that improved retrieval F1 by 10% and supported an estimated 4% productivity gain for enterprise users

BI Intern | (SQL, PL/SQL, MySQL, SSIS, DataStage, Cognos, Tableau, PowerBI) **Jul 2022 – Sep 2022**
WPG Holdings, Department of Business Intelligence
- Standardized procurement, inventory, and shipment data collection by enhancing 27 Cognos electronic forms, improving processing speed by 12% and reducing manual input errors by 8%
- Migrated and optimized 54 scheduled DataStage jobs to SSIS to support ERP data integration and regional consolidation, reducing data processing time by 13% while improving report loading performance and system stability
- Reduced reprocessing issues across monthly management reports, inventory monitoring, and sales tracking by 15%, supporting an estimated 50K in annual cost savings through lower maintenance effort

DATA SCIENCE PROJECTS

Transformer-Based Model for Video Segmentation in GUI domain | (Python) [GitHub]
- Built an end-to-end video segmentation pipeline on 2,000+ GUI interaction videos to model user action sequences from high-dimensional, unstructured visual data
- Fine-tuned temporal models, ASFormer, and designed a Swin Transformer with spatial attention to improve representation learning for high-resolution GUI environments
- Evaluated action-level classification performance, achieving a 38% accuracy improvement and demonstrating the potential of attention-based video analytics for automated interaction analysis and system optimization

Burn-in Board Dashboard Development | (MySQL, Microsoft SQL Server, Tableau) [GitHub]
- Reconciled data across MySQL SQL Server and Tableau extracts to resolve discrepancies and improve reporting reliability
- Rebuilt the data pipeline in Python and SQL to expand coverage and improve machine allocation accuracy by 16%
- Redesigned Tableau dashboards for cross team monitoring and faster pattern detection in burn in operations

Reinforcement Learning for Health Recommendation System | (Python, SQL) [GitHub]
- Developed a reinforcement learning–based recommendation system to personalize health activity suggestions using historical engagement signals and user context
- Balanced exploration and exploitation with Thompson Sampling, enabling adaptive selection under user response dynamics
- 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]
- Built an interactive dashboard over 60,000+ tornado events with pre aggregated metrics and interactive filtering
- Streamlined user experience through Streamlit interface with custom sidebar filters, real-time chart updates, and intuitive layout for seamless stakeholder ad-hoc 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]
- Structured and analyzed a high dimensional tabular dataset by staging data preparation and modeling steps to support repeatable and reliable premium prediction
- 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