Hao-Chun (Oscar) Shih

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

ACADEMICS

University of Michigan, Ann Arbor **Aug 2024 – May 2026**
Master of Data Science **GPA: 4.00**
Highlight Coursework: Data Mining, Database Management Systems, Information Visualization, Machine Learning

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

COMPUTER SKILLS

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

WORK EXPERIENCE

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

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

Facebook Data Management System | (MongoDB, Oracle, Java, JSON, SQL) [GitHub]
- Migrated 10,000+ records from Oracle's Facebook to MongoDB, ensuring seamless data transition
- Optimized 7 MongoDB queries, reducing data retrieval time by 25%
- Streamlined user data management, cutting storage footprint by 15% with MongoDB integration

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 different Parameter Efficient architecture, 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