Andy Huang

andyjhuang2639@gmail.com | linkedin.com/in/andy-jc-huang | github.com/anh1231

EXPERIENCE

San Diego Association of Governments (SANDAG) Jun 2025 – Present
Data Science Intern San Diego, CA

- Built a cloud data warehouse in Microsoft Fabric from scratch using a Lakehouse architecture to support scalable data ingestion, transformation, and storage
- Developed Dataflow Gen2 pipelines to migrate data into a structured, maintainable environment for analytics and reporting
- Designed ingestion and transformation workflows improving data accessibility and reliability across datasets
- Queried and validated large datasets using SQL Server, implementing reconciliation checks to ensure reporting accuracy
- Automated CSV processing, email data extraction, and report generation using Python, reducing manual workload by ~10 hours/week

PROJECTS

LLM Speculative Decoding on TPU (DFlash) Sep 2025 – Feb 2026

University of California, San Diego

- Ported diffusion-based speculative decoding from PyTorch/GPU to JAX/TPU within the vLLM inference framework
- Achieved 2.85x serving speedup (265 TPS) and up to 3.02x standalone speedup (773 TPS) on TPU v4/v5p
- Diagnosed and fixed sequence-length alignment bug, improving acceptance length from 2.49 to 4.75 and speedup from 1.30x to 2.85x
- Contributed production-ready code including model integration, proposer logic, benchmarking tools, tests, and CI
- **GitHub:** tpu-spec-decode

Brain-Computer Interface (BCI) Mouse via EEG Jan 2025 – Apr 2025

University of California, San Diego

- Developed a BCI-powered cursor control system using EEG signals and SSVEP classification for hands-free interaction
- Achieved 100% directional accuracy across 25 trials through optimized signal processing and classification pipelines
- Improved EEG data collection efficiency by 15% through enhancements in experimental setup and data acquisition

League of Legends Data Analysis & Baseline Model Jun 2023 – Jul 2023

University of California, San Diego

- Built a logistic regression model using gameplay statistics from 50,000+ matches to predict match outcomes
- Performed univariate and bivariate analysis to identify key features influencing outcomes
- Improved model accuracy by ~10% while analyzing bias and feature importance
- **Website:** Project Link

EDUCATION

University of California, San Diego Sep 2022 – Jun 2026
B.S. in Data Science San Diego, CA

SKILLS

Languages: Python, Java, SQL
Tools & Platforms: Microsoft Fabric, SQL Server, GitHub
Technologies: Machine Learning, Data Analysis, Data Visualization