Yuanrui (Jerry) Zhu

(510) 989-5064 | yuanruizhu2025@u.northwestern.edu | 1500 Sherman Ave, Evanston, IL 60201

EDUCATION

Northwestern University Sep. 2024 – Expected Dec. 2025

M.S. in Machine Learning and Data Science

University of California, Berkeley Sep. 2019 – Dec. 2023

B.A. in Computer Science & B.A. in Data Science (Double Majors with High Distinction Honor) *Overall GPA: 3.95/4.0*

SKILLS

Programming: Python, SQL, R, Java, C/C++, Pandas, scikit-learn, Matplotlib, Plotly, Pytorch, Spark, NLTK, OpenMP

Tools and Frameworks: Git/Github, Docker, AWS, Streamlit, CUDA, VS Code, RStudio, Excel, PowerPoint, Tableau

General: Data Mining, Deep Learning, A/B testing, Statistical Modeling, Data Visualization, Data Engineering, NLP

PROFESSIONAL EXPERIENCE

Baidu Jun. 2024 – Aug. 2024

Software Engineer Intern *Beijing, China*

- Developed a pipeline for generating randomly structured synthetic text, math formulas and tabular images to support end-to-end InternVL2 OCR model testing, utilizing Docker for consistent environment management
- LoRA fine-tuned InternVL2-8B on 1,500+ generated samples with GPUs, achieving a 1.3% reduction in character error rate
- Fine-tuned a pretrained Ernie-3.0 model using the PaddlePaddle framework on an internal dataset for multiclass news title classification and compressed the model for efficient online deployment via API

Honda Jan. 2023 – May 2023

Data Scientist Intern (Berkeley Data Science Discovery Program) *U.S. Remote*

- Applied Extract, Transform, Load (ETL) processes to parse raw vehicle Event Data Recorder (EDR) data, detected abnormal outliers, and stored the processed data in a PostgreSQL database for downstream applications
- Collaborated with domain experts to work on automated calculation of key dashboard features from the cleaned data
- Built a 3-level Streamlit BI dashboard to provide stakeholders with insights into vehicle trips, incorporating Google API for geographic functions and Plotly for interactive statistical visualizations, supporting data-driven business decision-making

JD.com (Jingdong) Jun. 2021 – Jul. 2021

Operational Data Analyst Intern *Beijing, China*

- Performed query optimization on existing SQL queries to enhance overall readability and reduced runtime by more than 10%
- Integrated data from diverse sources and set up real-time connections in Tableau to generate a comprehensive business report with insights into a new cloud product, contributing to a 4% increase in monthly profit margins

RESEARCH PROJECTS

Squishy Robotics: Machine learning for Wildfire Onset Detection and theDesignExchange Mar. 2023 – Present

Research Assistant, Berkeley Expert Systems Technologies Lab *Berkeley, CA*

- Conducted Text Analysis using Latent Dirichlet Allocation to analyze design method justification from human and GPT-3.5 generated reasoning. Co-authored a Design Research Society (DRS) conference paper https://doi.org/10.21606/drs.2024.956
- Enhanced task-specific segmentation Dice Coefficient by 6% by fine-tuning Meta's SAM model on wildfire images and masks
- Engineered prompts and developed programming solutions using GPT-4 API to extract key features from Corsican wildfire images with 83% accuracy, aimed at enhancing targeted wildfire management and response strategies

Human-Centered AI for Collaboration Sep. 2022 – Dec. 2023

Lead Research Assistant, Berkeley Operations and Behavioral Analytics Lab *Berkeley, CA*

- Coded and updated a Markov Decision Process (MDP) card game solver (30x runtime reduction for finding the most efficient solution) and an algorithm to calculate distance between players' card drawing sequences to assist team formation experiments
- Enhanced existing research outcomes by integrating counterfactual analysis using K-Means Clustering algorithm
- Received URAP summer research award under nomination https://research.berkeley.edu/urap-researchers/yuanrui-zhu/

Industry-level Movie Recommender System and User Study Assessment Sep. 2022 - May 2023

UC Berkeley Data Science Honors Student *Berkeley, CA*

- Utilized TF-IDF and Non-Negative Matrix Factorization for movie recommender systems, implementing content-based and collaborative filtering on public movie information and rating datasets with offline simulated evaluations
- Evaluated all studied systems by applying Kruskal-Wallis test and Tukey HSD to recorded ratings from 30 study participants
- Delivered an detailed analysis of user data and feedback to inform recommender system enhancements in a 15-page thesis