

EDUCATION

University of Pennsylvania M.S.E in Data Science	Philadelphia, USA 2023.09 - 2025.05
Peking University, Guanghua School of Management M.A. in Finance (Business Analytics track)	Beijing, China 2020.09 - 2022.06
Shanghai University of Finance and Economics B.A. in Management and Information System	Shanghai, China 2016.09 - 2020.06

WORKING EXPERIENCES

Meituan Technology Co., Ltd <i>ML Engineer Intern, Recommendation System</i>	2024.05 - 2024.08
<ul style="list-style-type: none"> Maintained and developed the data pipeline and utilized Spark and Hadoop to process data ETL Assisted with design of A/B test on the collaborate basis of products and CTR/RR, which result in 8.7% increase in revenue Utilized an offline RL algorithm w/ Diffusion policy to improve the performance of classical NCF algorithm 	
Sagehas Investment Co., Ltd <i>Machine learning researcher, Quantitative Research</i>	2022.04 - 2023.02
<ul style="list-style-type: none"> Utilized Autoregressive models to distill and denoise the trending and valuable factors from the financial time series with blanks and heavy noises Implemented geometric learning techniques to model market volatility on a latent manifold; This helped our team to build a strategy with 11.9% return rate in Q4 2022 on the hedge of USDX and US 10-year treasury 	
Tencent Technology Co., Ltd. <i>Data Scientist Intern, Interactive and Entertainment Group</i>	2021.01 - 2021.05
<ul style="list-style-type: none"> Made efficient SQL queries to interact with HDFS, and utilized PySpark to do cluster analysis Implemented A/B test and related inference on different user groups; Our new analytical tool helped improved 22.4% efficiency to recall users and convert non-active players to active ones Realized data visualization and interactive report via Echarts and Tableau. 	

RESEARCH EXPERIENCES

Dynamic Rank Adjustment in Diffusion Policies for Efficient and Flexible Training <i>RSS 2025 in Submission</i>	2024.09 - 2025.01
<ul style="list-style-type: none"> Assisted with training & deploying SAC and PPO to train expert policy; implemented a Low Rank Adaptation (LoRA) guided latent diffusion policy model over Gym based robotics platform which involves several different tasks Participated in combining the offline expert policy and online agent policy to do dataset aggregation imitation 	
Image editing based on Consistency Distillation from Pretrained Stable Diffusion Model <i>Research Project</i>	2024.11 - 2025.01
<ul style="list-style-type: none"> Implemented LoRA on pretrained distilled stable diffusion models to do sequential fine tuning; the performance is similar with SOTA image editing methods i.e. DiffusionCLIP and Stable Diffusion + PnP, whereas the speed increased 8x Generalized an inversion-free method to do image editing task inspired by the original paper; This involves the work of Control Net and a new perspective of noise scheduling 	
Classifier free guidance w/ Diffuser in Q-value Reinforcement Learning <i>Research Project</i>	2024.06 - 2024.08
<ul style="list-style-type: none"> Integrated the classifier free guidance into an offline diffusion policy RL model; Improved the performance on the by 7.2% percent than original baseline RL algorithms on Mujoco locomotion benchmarks 	
Large language models for financial reports analysis <i>Wharton Research Data Service</i>	2023.10 - 2024.03
<ul style="list-style-type: none"> Participated in the research related to a synthetic LLM implementation in RAG and in-context learning for entity recognition, clustering, and state change inference in procedural text; this outperformance the baseline FinBert+Llama on both precision rate and long-context information capturing 	

PERSONAL PROJECTS

Penn Cloud <i>CIS 5050 Software Systems, Course Project</i>	2025.03 - 2025.04
<ul style="list-style-type: none"> Developed a distributed cloud platform offering a webmail service and storage service (similar to Gmail and Google Drive) using C++; utilized Protobuf for data serialization and gRPC for inter-component communication Realized a load-balanced frontend system that pings a central load balancer for health monitoring and provides user interfaces for email and file storage, along with an admin console to view backend status and BigTable key-value pairs. 	

- Distributed Search Engine**
CIS 5550 Internet & Web Services(Audited), Course Project

2024.11 - 2024.12

 - Built crawler based on **Apache Storm** which was hosted on **EC2**, and stored the content in **S3** bucket
 - Utilized **MapReduce** to build indexer, which segmented the documents into indices and stored the indices in **DynamoDB**
 - Built **PageRank** based on **Spark** and run on **Elastic MapReduce** clusters for searching
- Penn Treehole**
Fullstack Development Project

2024.02 - 2024.03

 - A college-based anonymous forum base on **DynamoDB – Express – React - Node.js** stack
 - Hosted the website on AWS **EC2** and utilized **API Gateway** to realize the load balancing

PERSONAL QUALITIES & QUALIFICATIONS

Programming Language	Python, C++, SQL, Cuda, Java, JavaScript, Node.js
Research Skills	Hugging Face, Pytorch, Pandas, Git
AWS Skills	EC2, Lambda, RDS, S3, DynamoDB
Tools & Platforms	Linux, Spark, Hadoop, Nginx, Docker, Git, Hive, Node.js, Redis