

Benny Xinhao Jiang

bennyjiangxinhao@gmail.com | 510-207-4905 | linkedin.com/in/benny-jiang | github.com/Bennyoooo

EDUCATION

Carnegie Mellon University

Master of Computational Data Science, Analytics Concentration

- Relevant Coursework: Deep Learning, Advanced Natural Language Processing, Machine Learning.

Pittsburgh, PA

May 2021 - May 2023

University of California, Berkeley

Bachelor of Arts in Computer Science

Bachelor of Arts in Data Science with domain emphasis in Applied Mathematics & Modeling

- Relevant Coursework: Deep Reinforcement Learning, Parallel Programming, Database Systems, Operating Systems.

Berkeley, CA

Aug 2017 - May 2021

SKILLS

Data Science: PyTorch, Scikit-learn, NumPy, SciPy, Pandas, Seaborn.

Tools & Technologies: Spark(MapReduce, SparkML), AWS&Azure(machine learning), MySQL, MongoDB, Linux.

Programming Languages: C++, Python, Java, Javascript(AngularJS), Swift.

PROFESSIONAL EXPERIENCES

Tencent

Software Engineer Intern, Financial Technologies

- Collaborated with a team of 15 to improve the transaction and user management of digital currency platforms.
- Constructed an RPC server using C++ that enabled customers to forge secure payment channels with merchants.
- Upgraded an RPC framework that supported multi-process reverse proxy, and load balancing of servers.

Shenzhen, China

May 2021 - Aug 2021

Hulu

Software Developer Intern

- Built a troubleshooting server with Flask that reduced the complicated procedures of locating errors on multiple micro-services into one simple searching action.
- Constructed an integration platform that monitored the status of seven micro-services, saving time of engineers across five teams for inspecting and debugging the servers.
- Led a project team of 3 to create a Chrome extension with JavaScript that let users collect and share video clips in TV shows and won a special award in the internal corporate hackathon.

Santa Monica, CA

Jun 2020 - Aug 2020

BizSeer

Research Data Scientist Intern

- Researched a framework to diagnose the root causes of slow queries in large-scale cloud databases with ML technologies.
- Proposed a clustering algorithm combined with case-based reasoning models to create an active learning pipeline that queried the labels of representative data points from database experts, efficiently easing labeling overhead.
- Developed a real-time anomaly detector that extracted and categorized anomalies on time series, with over 97 F1 score.
- Co-authored and published a paper: "Diagnosing root causes of intermittent slow queries in cloud databases" in Proceedings of the VLDB Endowment 13, no. 8, April 2020.

Beijing, China

Jun 2019 - Aug 2019

PROJECTS

Multi-lingual Factual Knowledge Probing

- Evaluated large pre-trained language models as factual knowledge bases in 17 different languages by prompt engineering.
- Conducted comprehensive literature review, experiments, and error analysis on different methods of multilingual knowledge probing with pre-trained language models.
- Devised Cross-lingual Case Analogy prompting that improved transfer learning of knowledge probing by over 5%.

Sept 2021- Dec 2021

Voice Is All You Need

- Proposed a GAN-based framework to reconstruct face images from the speakers' voices.
- Designed an evaluation method using pre-trained FaceNet to assess the quality and fidelity of generated face images.
- Preprocessed and cleaned the training data by cropping face frame from noisy image datasets.
- Built and trained the CNN based voice embedding network combined with a GAN based framework that can generate sensible face images reconstructing important facial features.

Sept 2021- Dec 2021

Music Generation with Transformer Language Models

- Constructed a Transformer language model, trained on two music domains: classical songs and pop songs, that generated new music in two different styles.
- Analyzed generated songs by pitch range and pitch transition matrix to compare music in different style quantitatively.
- Conducted online surveys, demonstrating that the generated songs have clear structures and aesthetic richness.

Oct 2021- Dec 2021

Multi-task Deep Reinforcement Learning with Parameter Superposition

- Initiated a research project to resolve catastrophic forgetting in multi-task deep RL with parameter superposition.
- Demonstrated through experiments that deep RL with parameter superposition increases its capacity of overcoming catastrophic forgetting by over 30% on average compared to baseline methods.
- Implemented various deep RL models with PyTorch, including Policy-Gradient, DQN, model-based RL, Actor-Critic; trained and tested different models in MetaWorld environment in experiments.

Aug 2020- Nov 2020