

Jianheng (Jaden) Hou

Phone Number: (213)-204-0607 | Email: jianhenh@gmail.com | Address: 312 Gates Dr. Milpitas, CA95035

[LinkedIn Profile](#) | [Portfolio Profile](#) | [ML-backed FinTechApp](#)

Professional Experience

Senior Machine Learning Engineer @ Tesla

June 2023 - Present in Fremont, CA

- Scoped out the first **one-year** project and led effort with 4 engineers, 1 PM, and 10 domain experts to build out a **nearline multi-modal recommendation system and a multi-task classification system** [CLIP LoRA fine-tuning; MMoE+Multi-Task Head] involving 2k+ overlapping labels. Sealed the system to all global Tesla Gigafactories, **saving 3M+/year in labor costs**;
- Scoped out the second **half-year multi-team project** and led 2 engineers, and 1 intern to build a chat Agent, as a cooperative-level business workflow automation solution, that routes to **GraphRAG and serveral language-to-query tools**, which provides **local qualitative, global statistical, and global summative answers to user's requests** and **saved \$5M/year labor cost**;
- Built **Active Learning & Human Review System** for high-cardinality Taxonomy Compilation and Labeling workflow via Content Semantic Visualization and Human-in-the-Loop.
- Designed metrics for the team and built **dashboards, reports, and alerts** to deliver actionable insights for leadership.
- Created and maintained **20+ Airflow automation pipelines** in the Kubernetes environment.
- Built up data science portfolio including an all-purpose **ML-based causal analysis framework** and **Search Optimization**
- Recognized as one of the 3 highest-impact employees in the org for 2024

Senior Data Scientist, Algorithm @ Comcast Corp

Feb 2023 – June 2023 in Philadelphia, PA

- Served in a rotation program in the Enterprise Business Intelligence strategic analytics team and delivered analytics/AI solutions for stakeholders in customer service, product, growth, workforce, and applied data science teams respectively.
- Cancellation fee short/long term recommendation by A/B test and business analytics

ML Data Scientist @ Alpha Edison (Venture Capital)

May 2019 – Feb 2023 in Los Angeles, CA

End-to-end Text Mining: Developed a [Topic Based Sentiment Analysis System](#) to recognize medical industry issues and latent opportunities & enable stakeholders to successfully target high-potential MED startups and make investment decisions

- Conducted **web scraping with Scrapy**, data integration on **AWS S3**, and NLP preprocessing with **PySpark**;
- Designed labeling processes for training data through **Mechanical Turk** and optimized efficiency(10X+) under a limited budget;
- Built **NLP models (CNN+Bi-LSTM+Attention)** with **Tensorflow** to predict topics and sentiment of patient's posts (10M rows);
- Deployed **data pipeline automation, modeling training/serving, monitoring** on Databricks
- Delivered a dashboard (SQL-based) with search panels on **Databricks** to report and alert metrics, charts, and summaries.

[Time-series Anomaly Detection & ML-based Factor Analysis](#): An ML-based analytics framework that tracks the unexpected consumption trends across all US E-commerce platforms before its manifestation at scale and identifies consumer segments that behave differently and associated quantitative causal factors.

- Developed **time-series pipeline** to track anomaly consumption trends at sector level and rank top movers with indicators;
- Used **XGBoost Model and SHAP** to identify top demographic features as the causal factors of unexpected consumption trends;
- Developed dashboards integrating metrics, diagrams and NLP summaries for non-technical partners to use on a regular basis;
- Causal Inference Effect of consumer political affiliation on their variety-seeking behavior via **propensity score matching**

Research Assitant @ Information Sciences Institute – Knowledge Graph Center

May 2018 - Dec 2018 in Los Angeles, CA

[Text-enabled Humanitarian Operations in Real-time](#)

- Conducted distributed data processing on heterogeneous data and event entity resolution for knowledge graph;
- Developed an online clustering algorithm on tweets that dynamically identify event situations to support rescue teams;

Education

University of Southern California

Los Angeles, CA

Master in Data Informatics (Applied Data Science)

Sep 2017 - Dec 2019

- Machine Learning, Deep Learning, NLP, Data Mining, Knowledge Graph, Statistics, Causal Inference, Experiment Design

Skills

- **Machine Learning and Data Science:** pandas, Tensorflow, sklearn, Keras, LangChain, Huggingface, LlamaIndex, OpenAI, NLTK, OpenCV, Scipy, MLFlow, SageMaker, SQL, Tableau,
- **Distributed Systems, Platforms and Database:** Kubernetes, AWS(Amplify, MWAA, DBS, Lambda), Kafka, Flink, Airflow, Databricks, Github, Spark, Hadoop, Flask, Redshift, Neo4, MongoDB, MySQL