Linsu Han (ltbd78.github.io/portfolio)

(317) 515-1299 • New York • linsu.han@columbia.edu • github.com/ltbd78 • linkedin.com/in/linsuhan

EXPERIENCE

Senior Software Engineer, Machine Learning Oct 2023 - Present

Walmart | Bentonville, AR

Spearheading Walmart's new Retail Intelligence team's GenAI research and pipelining solutions.

Key Accomplishments

- Deployed a web app demoing Image2Text and VLM models (ViT+GPT, BLIP, GIT, moondream, DeepSeek) used for labeling Walmart's internal database of images. (Tech Stack: FastAPI, Streamlit, Hugging Face, PyTorch, GCP, Docker)
- Developed an internal document Retrieval Augmented Generation (RAG) pipeline proof of concept. (Tech Stack: LangChain, Llama2, SentenceBERT, Chroma Vector DB, PyTorch)
- Lead internal DS/MLE training sessions on how to build an end-to-end training-to-deployment deep learning pipeline leveraging a custom-built GPT model from scratch (Tech Stack: PyTorch, Kubeflow, GCP, Vertex AI, Docker)

Senior Data Scientist Sept 2021 - Aug 2023

Capital One | New York, NY

Technical contributor for a recommendation system that leverages a Contextual Multi-Armed Bandit to prescribe personalized contact strategies for customers with delinquent accounts.

Key Accomplishments

- Received a Capital One Recognition Award for redesigning model pipeline to be more modular and customizable, reducing model development costs from weeks to hours and improving performance. Model was implemented in Capital One's Collections Contact Optimization team's infrastructure and successfully launched into production.
- Built an interactive evaluation framework that can rapidly compare models and generate interactive plots of various performance metrics including ROC, precision recall curves, confusion matrices, top-K feature importance, Brier score losses, and probability calibration curves.
- Collaborated with the MLE team to adapt old pipelines to work with Kubeflow, allowing parallelized training and evaluation of ML models.

Deep Learning Research Scientist Jan 2020 - Sept 2021

Boston Fusion | Lexington, MA

Juggled several SBIR and DoD research projects. Technical mentor for new hires. Gave company-wide technical presentations on topics such as Reinforcement Learning and Graph Convolutional Networks.

Key Accomplishments

- Devised a reinforcement learning OpenAI-Gym-like framework for ROS (Robotics Operating System); trained a Dueling Double DQN with Prioritized Experience Replay agent to successfully navigate a UAV to a target destination while simultaneously avoiding moving obstacles.
- Improved acoustic classification of spectrograms and MFCCs with ResNet architecture increasing 5-fold cross validation accuracy by 5%; upgraded the framework from single-label to multi-label classification; developed a Convolutional Autoencoder to detect anomalous acoustic signals; developed a Siamese Network to compare in-distribution data with out-of-distribution new acoustic signals.
- Prototyped a novel recommendation system that combines the classic Graph Convolutional Network and a Contextual Multi-Armed Bandit framework to enable inferences on graphs; extended research paper from node-level to graph-level.
- Lead initiative on redesigning/rewriting ML pipelines of 2 major projects into clean and efficient pipelines; merged multiple training threads to a single loop, eliminating redundant data queries and reducing memory load by 4x; parallelized training process on multiple GPUs significantly speeding up training process by a factor of 4.

Data Science Intern Jun 2019 - Jul 2019

Graphen, Inc. | New York City, NY

- Built a web crawler with customizable parameters that automates scraping news articles and archiving them in a database.
- Applied natural language processing techniques to extract key features from articles for analysis (i.e. NER).
- Coded a script to analyze tweets in real-time and predict sentiment scores.

PROJECTS (github.com/ltbd78)

gpt	| GPT from scratch in PyTorch and scaled model deployment via Kubeflow on GCP Vertex AI.
Lost Ark Omnibot	| A human-like pixel bot that plays Lost Ark MMORPG [youtu.be/Zntd6JhbVCo]
cae	| Convolutional Autoencoder used for image reconstruction or anomaly detection.
RL	| Flexible RL framework that allows easy mix and matching of different agents and environments.
facial-feature-detection	| Flask hosted, real-time face detection & feature classification; ResNet model trained on CelebA dataset.
pairtrader	| Classic pair trading algorithm with backtesting on stock and crypto markets.
algorithms	| Collection of random algorithms (Adaboost, Expectation Maximization, K-Means).

EDUCATION

Columbia University | Master's in **Statistics** (GPA: 3.90/4.00) 2018 - 2020

Purdue University | Bachelor's in **Mathematics** & **Statistics** (GPA: 3.90/4.00) 2014 - 2018

- *Magna Cum Laude, Presidential Scholarship, Leath Scholarship, Dean's List, Semester's Honors*

SKILLS

Python ⊇ {jupyter, pytorch, tensorflow, sklearn, kubeflow, spacy, numpy, scipy, pandas, matplotlib, opencv, pymysql, flask, ...}
Machine Learning ⊇ {boosting, bagging, GLMs, decision trees, SVM, PCA, naive bayes, KNN, Gaussian Mixture Models, …}
Deep Learning ⊇ {transformers, NLP, CNN, reinforcement learning, autoencoders, gen ai, siamese nets, graph neural nets, ...}
Git | Google Cloud | Docker | SQL | React | Javascript | HTML | CSS | Bash | R | Java | SAS | CUDA | ROS | MATLAB