Vishal Tien

Website | LinkedIn | GitHub | Email

SUMMARY

Passionate machine learning engineer with >2 years of experience spanning both research and industry, specializing in natural language processing and knowledge graphs. Domain expertise in healthcare and looking to grow my skills in other industries

SKILLS

Technologies: PyTorch, TensorFlow, Scikit-learn, AWS, Flask, React, Apache Spark, NoSQL, Neo4j, Docker, Kubernetes
Programming Languages: Python, Ruby, Java, JavaScript, R, MATLAB, SQL, HTML, CSS

WORK EXPERIENCE

Roivant Sciences, New York, NY — July 2021 – Present
Tech Rotational Analyst

- *Software / MLOps Engineer*
  - Led development of organization's first MLOps tools, increasing model development / deployment velocity, reducing costs by 40%, and promoting CI/CD, infrastructure-as-code (Chef, Terraform), and containerized development best practices
  - Orchestrated distributed ETL pipelines for production data warehouse that streamlines sales analysis and customer targeting
- *AI for Knowledge Discovery Team - NLP Scientist*
  - Developed tool that combines rule-based pattern matching and a fine-tuned T5 language model to assist researchers in retrieving relevant spans of text from millions of documents stored in NoSQL database, resulting in the reduction of manual curation efforts by >50% for information retrieval tasks
  - Tech lead for a knowledge graph based chatbot application with a Neo4J backend powered by custom fine-tuning GPT-3
- *Digital Innovator*
  - Led the design and development of ML product that analyzes Sumitovant's corporate reputation using information extracted from the web, enabling recommendation of marketing actions to the VP of Communications in a Flask application
  - Built and deployed NLP deep learning models (i.e BART) using HuggingFace and AWS SageMaker for topic modeling and document classification trained with zero-shot learning approaches and human-in-the-loop techniques, enabling automation in a data scarce environment
  - Integrated ElasticSearch into backend to increase speed of information retrieval and developed an interactive d3.js frontend visualization for React application that streamlines manual processes in research scientist drug discovery workflow
- *AI for Knowledge Discovery Team – Knowledge Graph Engineer*
  - Built a biomedical knowledge graph from disparate data sources (1M+ nodes and 4M+ edges) and designed ontology using RDF to power a semantic search engine and enable previously unanswerable drug discovery questions to be addressed
  - Implemented graph-based algorithms / GNNs to uncover insights from knowledge graph, such as disease-disease similarity
  - Trained a question answering information retrieval system by implementing an English to SPARQL generative language model, enabling the creation of a natural language interface that democratized the knowledge graph across the organization

Tsui Lab at Children's Hospital of Philadelphia, Philadelphia, PA — Dec. 2020 – July 2021
AI Researcher

- Constructed a CNN-LSTM model in PyTorch that predicted presence of life-threatening cardiac condition from multi-channel time-series waveform data, ultimately improving performance of current state of the art hospital solution by 77%

Merck & Co., Inc., Branchburg, NJ — Jun. 2020 – Aug. 2020
Data Science Intern – IT Emerging Talent Program

- Built machine learning classification model with > 99% accuracy in Python to uncover relationships between large structured / unstructured datasets without data dictionaries, improving recall of previously best performing model by ~50%

EDUCATION

University of Pennsylvania, Philadelphia, PA — August 2016 - May 2021
BSE in Bioengineering | Minors: Mathematics and Engineering Entrepreneurship (May 2020) | GPA: 3.8/4.0
MSE in Systems Engineering | Concentration in Data Science (May 2021) | GPA: 3.94/4.0
Awards: Senior Design Award, Rothberg Catalyzer Award, Publication, BMES Student Design and Research Award, Dean's List

PROJECTS

Podcast Trailer Generator

- Implemented research paper approach to automate trailer generation from podcast audio and deployed using Docker and FastAPI

A Comparison of Abstractive and Extractive Text Summarization using Question Answering Benchmarks

- Deep learning research paper that contributes to the field as part of final project for deep learning graduate course

Neural Network Implementation From Scratch

- Constructed neural network from scratch using pure python and numpy and compared performance to PyTorch network