

Vishal Tien

[Website](#) | [LinkedIn](#) | [GitHub](#) | [Email](#)

SUMMARY

Passionate machine learning engineer with >2 years of experience in ML spanning both research and industry, specializing in natural language processing. Domain expertise in pharma / healthcare and looking to grow my skills in the tech industry

SKILLS

Technologies: PyTorch, HuggingFace, Scikit-learn, AWS, Flask, React, Spark, NoSQL, Neo4j, Docker

Programming Languages: Python, C++, R, MATLAB, SQL, JavaScript, HTML, CSS

WORK EXPERIENCE

Roivant Sciences, New York, NY

July 2021 – Present

Tech Rotational Analyst

- *Sumitovant, 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 documents (i.e those mentioning drug side effects) from millions stored in NoSQL database
 - Tech lead for a knowledge graph based chatbot application powered by fine-tuning GPT-3 with hand-built Neo4j queries
- *Sumitovant, 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
- *Sumitovant, AI for Knowledge Discovery Team – Knowledge Graph / NLP Scientist*
 - Built a biomedical knowledge graph (1M+ nodes and 4M+ edges) using Python and Turtle from disparate data sources 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 (graph query language) deep learning generative language model, enabling the creation of a natural language interface that allows non-technical users to benefit from knowledge base, drastically increasing the impact it can have

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 clinical patient 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%
- Developed API wrapper written in R to allow data scientists to interact with a core product's API through easy-to-use functions

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

[An Evaluation of Abstractive and Extractive Deep Learning NLP Text Summarization Techniques](#)

- Final project for Principles of Deep Learning graduate course – required contributing to the deep learning research field

[Implementing Neural Network From Scratch](#)

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

[Predicting Inpatient Length of Stay at Hospitals \(TowardsDataScience Blog Post\)](#)

- Applied ML to > 2 million rows of patient data to uncover clinically relevant insights and predict patient length of stay

[NYC Restaurant Finder](#)

- Developed full-stack React / Node.js web application ([demo](#)) providing advanced restaurant search capabilities