

HAOYU (VINCENT) TIAN

(857) 320-8907 | vinnyt@mit.edu | [LinkedIn](#) | [Portfolio](#) | Australia Citizen (E3)

EDUCATION

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Cambridge, MA

Candidate for Master of Business Analytics, Operations Research Center, August 2024, GPA: 5.00/5.00 2023 - Present

- **Deep Learning Project:** Utilized Amazon product images to generate Google Gemini image captions, constructed 3K DALL-E 2 images and fine-tuned VGG19, Resnet50 and EffNet architectures, improved baseline error rate by 96% (*Python*)
- **Leadership: VP of Blockchain Club**, organized mixer event with MIT, Harvard Business School and Harvard Law School

UNIVERSITY OF MELBOURNE

Melbourne, Australia

Bachelor of Science in Mathematics and Statistics, GPA: 3.92/4.00

2017 - 2020

- **Awards:** Study Abroad Scholarship in Statistics - University College London (Spring 2019)
- Transferred from the University of Western Australia (Fall 2019)

TECHNICAL SKILLS

- **Languages:** Python (NumPy, Pandas, scikit-learn, PyTorch, TensorFlow, Keras), R, SQL, Julia/JuMP (Gurobi), Git, Java
- **Tools:** Snowflake, GCP (BigQuery, VertexAI), AWS (SageMaker, S3), MS SQL Server, Azure Data Factory, PowerBI
- **Skills:** Statistical/Data Analysis, Data Visualization, ETL, Clustering, LLMs, Data Quality Assurance, A/B testing, MLOps

PROFESSIONAL EXPERIENCE

MACY'S | MIT Capstone

New York City, NY

Data Scientist

Feb 2024 - Present

- Performing exploratory data analysis and constructing ETL pipelines on 240M rows of transaction, loyalty and product data
- Building customer price brand models and refining recommender systems for personalized e-commerce/in-store experiences
- Presenting to non-technical audiences, discussing product use cases with business stakeholders and directors

QUANTIUM

Melbourne, Australia

Data Scientist, Product Analytics

Jan 2022 – Jun 2023

- Leveraged OOP to integrate customer segmentations and reports into analytical testing process, capturing \$42M of revenue
- Pioneered customer engagement metrics using Q.Scan Amplitude data and built Decision Tree, Random Forest and XGBoost models in R for churn identification; presented to product executives
- Collaborated with Product Managers and Software Engineers to address client inquiries, plan tasks and facilitate debugging
- Mentored 10 Associate Data Scientists team members in Python, Snowflake, SQL, Git over one year

Associate Data Scientist, Health & Government

Feb 2021 – Dec 2021

- Elevated XGBoost models by 2% in Python, introducing the state government's first daily population estimation method, enhancing risk planning across all suburban areas during Covid-19
- Led a 5 person PowerBI workstream on a \$400K project to develop an end-to-end tool that analyzes trends in education and health, aiding in identification of at-risk communities; led to a \$2M contract extension
- Engineered automated decision support tool pipeline (Microsoft Power Automate) used in daily meetings by Emergency Management Commissioner and Victorian Chief Medical Officer, reduced weekly labor by 14 hours

FORTESCUE METALS GROUP

Perth, Australia

Data Scientist Intern, Machine Learning & AI team

Nov 2020 – Feb 2021

- Created Random Forest, LightGBM and Multi-Layer Perceptron models in Python to complement existing manual testing process, boosted oil quality prediction efficiency for technicians
- Implemented MLOps solutions, deploying a Random Forest model on AWS using Lambda and Step Functions
- Conducted data preprocessing and exploratory analysis of oil testing data in AWS Sagemaker, gained experience using S3

PROJECTS

WEB3 AI STARTUP

Cambridge, MA

Featured as an innovative AIxWeb3 project on the largest Web3 platform in the Asia-Pacific region Jan 2024 - Present

Co-Founder & AI Researcher

- Leading a team of 5 to build the first end-to-end AI-powered Web3 analytics tool that scrapes and clusters text data from Telegram, Discord and Twitter to generate personalized market and sentiment analysis insights using open-source LLMs
- 1/17 finalists selected from 238 teams at Edge Matrix Computing (EMC) DeAI Hackathon; Presented to audience of 30 consisting of EMC CEO, AI Lead of EMC Lab, AI and ML Researchers, Professors, VC and Web3 founders

ADDITIONAL INFORMATION

- Mandarin (fluent), Interests: Novice Film Reviewer, Dedicated NBA/Soccer fan (English Premier League)