

CHUNG-YEH(OLIVER) YANG

(646) 992-7153 | cy2816@columbia.edu | linkedin.com/in/olivery0307 | https://github.com/Olivery0307

EDUCATION

Columbia University

New York, NY

M.S. in Data Science

Dec 2026(Expected)

Relevant Coursework: Exploratory data analysis, Computer Vision, Operational Logistics, Machine Learning, Poisson Process

Boston University

Boston, MA

B.S. in Data Science, GPA: 3.8/4.0

Sep 2020 - Dec 2024

Major in Data Science, Minor in Business Administration

Coursework: Bayesian Statistics, Computer Systems, Database Design, Data Structure Algorithms, Deep Learning, Hypothesis Testing, Linear Algebra, Machine Learning, Multivariable Calculus

WORK EXPERIENCE

PalAI

New York, NY

AI/ML Intern

Sep 2025 – Present

- Engineered and prototyped a core ML and AI-driven college application and essay feedback platform
- Designed a data pipeline to standardize and augment 100+ essay/feedback pairs and **fine-tuning** the model with the **OpenAI API**
- Collaborated with cross-functional backend and frontend engineers to **deploy ML models** into a production-level user experience

Boston University Questrom School of Business

Boston, MA

Research Assistant in Applied AI in Business

Aug 2023 - Jan 2025

- Collaborated with **5+ industry researchers** to design and implement four AI solutions tailored to specific business challenges
- Analyzed over 50,000 user queries to measure the impact of query characteristics on **RAG generation** and user behaviors
- Engineered a pipeline that uses an **LLM** to label artwork genres from unsupervised clusters, achieving **zero-shot topic modeling**

BU Spark!

Boston, MA

Data Science Technical Project Manager

Sep 2024 - Dec 2024

- Managed six data analysis and visualization projects, coordinating with clients and facilitating student communication
- Provided technical assistance and leadership on data visualization projects leveraging **Python, Tableau, and Google BigQuery**
- Configured dataset and software setup for Notion and **GitHub**, ensuring teams had resources for project milestones

GIGA RESET

Shanghai, CN

Data Science Summer Intern

Jun 2024 - Aug 2024

- Designed a data processing pipeline for the RESET Air project with Python, reducing manual workload by **80%**
- Automated creation of scorecards by importing **Python** libraries, optimizing slide generation and **data visualization**

PROJECTS

YouBike Prediction Service

Jun 2025 - Present

- Developed a serverless data pipeline on **AWS (Lambda, S3)** to collect and process real-time YouBike API data
- Engineered time-series features to train a **LightGBM** model, achieving over 90% recall in predicting critical station states
- Deployed and dockerized and prediction pipeline using AWS Lambda and ECR for automated, live inference

Deep Neural Networks Fusion Approach for VQA Verification

Oct 2024 - Dec 2024

- Built a custom preprocessing pipeline with **PyTorch** to balance and label correct and incorrect answers for effective training
- Constructed a fusion-based model combining image and text features to classify answers in Visual Question Answering
- Achieved 87.07% validation accuracy using **attention mechanisms**, showcasing strong generalization of model

Multi-Modal Topic Modeling for Artworks

Jan 2024 - Jul 2024

- Established a zero-shot topic modeling pipeline to identify and label thematic genres in a dataset of unlabeled artworks
- Implemented clustering utilizing **CLIP embeddings** and applied HDBSCAN to identify 15 distinct thematic clusters
- Pioneered a novel labeling technique by prompting a LLM with images to generate accurate, human-readable genre descriptions

LEADERSHIP EXPERIENCE

Taiwanese Overseas Students Association at Boston University

Boston, MA

President

Apr 2023 - May 2024

- Organized events and promote Taiwanese culture for over 300 Taiwanese students and general members in Boston University
- Coordinated and distributed work to 20 executive board members, ensuring efficient and event planning and execution

SKILLS

Programming Languages: Python, SQL, R, Rust

ML Libraries & Frameworks: PyTorch, LLM, OpenAI API, LangChain, Transformers, Scikit-Learn

Platforms & Developer Tools: AWS, Docker, Git, Google BigQuery, Tableau, Redis