# Erchi Zhang

**Email: ez806@nyu.edu | Tel: 646-906-7909 | Website: https://archertakesitez.github.io/**

## SKILLS

**Programming Languages:** Python, SQL, R, Java, HTML, CSS, JavaScript, TypeScript, Shell, Elixir
**Expertise:** Exploratory Data Analysis, Data Pipelines, Data Visualization, Distributed Systems, Spark, Hadoop, Databases, A/B Testing, Linux, Git, Machine Learning, LLMs, NLP, Generative AI, Computer Vision, NumPy, Pandas, Azure, AWS

## EDUCATION

**New York University,** New York, NY — Expected May. 2025
*Master of Science in Data Science* — GPA: 3.762/4.0
Coursework: Programming for Data Science | Optimization and Computational Linear Algebra | Big Data | Machine Learning

**Brandeis University,** Waltham, MA — Aug. 2019 – May. 2023
*Bachelor of Science in Computer Science, Bachelor of Arts in Economics, Minor in Mathematics* — GPA: 3.749/4.0
Coursework: Data Structures | Data Management | Deep Learning | Data Mining | Natural Language Processing | Algorithms

## PUBLICATION

*GraphBERT: Bridging Graph and Text for Malicious Behavior Detection on Social Media*. Published by ICDM-2022.
https://archertakesitez.github.io/static/assets/papers/GraphBERT.pdf

*Fair Graph Representation Learning via Diverse Mixture-of-Experts*. Published by The ACM Web Conference 2023.
https://archertakesitez.github.io/static/assets/papers/Representation_learning.pdf

## PROFESSIONAL EXPERIENCE

**Kineviz Inc.** — San Francisco, CA (remote)
*Data Analytics Intern* — Jun. 2024 – Aug. 2024

- Incorporated **LLMs** (**Large Language Models**) into a chat GUI via **Python** and **JavaScript**
- Developed a **Cypher** query parser with ANTLR in Python, incorporating Named Entity Recognition (**NER**) and **vector search** to validate and enhance Cypher queries, enabling advanced search capabilities in **Neo4j** graph databases
- Debugged and improved Kineviz's graph visualization tool (GraphXR) using **Azure** virtual machines

**Liangyouyinli Technology Co., Ltd** — Beijing, China
*Data Scientist Intern* — May. 2023 – Aug. 2023

- Utilized **PyTorch** to implement convolutional neural networks (**CNN**) for stock price prediction in China's stock market
- Through **backtesting**, evaluated our models by checking whether the selected top five stocks made gains in a designated time frame (3 or 5 days), concluding that the models predict with an accuracy of approximately 75%

**YUSUR Technology Co., Ltd** — Beijing, China
*Software Development Engineer Intern* — May. 2021 – Aug. 2021

- Queried datasets from **SQL** databases (**MySQL**, **PostgreSQL**) and **NoSQL** database (**MongoDB**)
- Utilized **Apache Spark** and implemented **Shell** and **Python** scripts to process the Iris dataset on **distributed and parallel systems**, benchmarking the performance of our self-developed Kernel Processing Unit (KPU)
- Automated **data pipelines** in **Python** to preprocess and clean data using **Prefect** and **Mage**
- Visualized our findings with **D3.js**, **Matplotlib**, and **Seaborn**, and made presentations using these data visualizations

## ACADEMIC PROJECTS

**Deck-to-CPT: AI-Driven Reimbursement Code Discovery for HealthTech Start-Ups** — Fall 2024

- Built a **Python**-based web application with **HTML/CSS** for processing PDF pitch decks and returning relevant Current Procedural Terminology (CPT) codes using **AI**
- Automated **web scraping** to retrieve detailed CPT code information from healthcare websites
- Applied Named Entity Recognition (**NER**) to extract key information from PDFs and utilized Retrieval-Augmented Generation (**RAG**) for accurate CPT code recommendations

**Fixplainer: Failure Explainer for Multiple Object Tracking (MOT)** — Spring 2024

- Created a **GUI** tool that can extract features and then generate various **SHAP** explanation plots for the objects in a multiple object tracking (**MOT**) task video frame, elucidating why the objects are successfully or unsuccessfully tracked
- Applied **YOLOv8** and **BoT-SORT** as object detection and object tracking tools on video datasets to create training sets