YUFAN ZHANG

914-720-6892 | yz2894@cornell.edu | New York, NY | GitHub | LinkedIn | Website

EDUCATION

Cornell Tech (Cornell University), New York, NY — May 2025
Master of Science in Information Systems – Connective Media Concentration

Duke Kunshan University / Duke University Dual Degree Program, Suzhou, China — May 2023
Bachelor of Science in Data Science | GPA: 3.7/4.0

TECHNICAL SKILLS

Coding Languages:	Python, Java, HTML/CSS/JavaScript, C++, SQL
Data Science & Machine Learning:	Pandas, NumPy, Scikit-Learn, PyTorch, TensorFlow Extended
Web Development:	React.js, Express.js, Node.js, Django, MongoDB, PostgreSQL, MySQL, Socket.io
Miscellaneous Skills:	Git, Docker, Jira, AirFlow, Linux, Conda

EXPERIENCE

eBay Inc., Product Manager, Internship (Cloud Data & Storage), Shanghai, China — Mar 2023 – Jun 2023

- Championed the **data streaming** platform, leveraging knowledge in **Kafka** and **Flink** to enhance platform adoption and performance
- Facilitated successful migration of **3 high-impact data messaging** use cases to a new messaging platform with enhanced **HA/DR** capabilities by orchestrating user meetings, evaluating dependency requirements, and partnering with the engineering team
- Boosted customer adoption of the new data governance platform, Data Lake Studio, by **24%** by leveraging **customized MkDocs** to maintain detailed product **documentation**, including user guides, API documentation, and technical specifications
- Organized BrownBag session that elevated customer awareness of the new **SQL-as-Stream** feature, resulting in **10 new onboardings**

Duke Kunshan University, Research Data Engineer (Data Science Research Center), Kunshan, China — Jul 2022 – Sep 2022

- Designed and implemented an automated Ethereum data ingestion pipeline using **Apache Airflow**, handling over **2 million** transaction records and updating data in near-real-time, which reduced manual intervention by **80%** and achieved **3x** accelerated research iterations
- Maintained a **PostgreSQL** instance on **Google Cloud SQL**, with optimized database schema for data retrieval with **75ms** average response time
- Implemented graph-based analytics using **NetworkX**, **Pandas**, and **Raphtory** for cross-sectional comparison of decentralization levels in DeFi protocols, resulting in a **first-authored** research framework at the *2023 Computing Conference* [**Paper**], which has received **3 citations**
- Leveraged an interactive dashboard using **Plotly Dash** to visualize the evolution of decentralization levels of each DeFi protocols

PROJECTS

miniTorch, (Python) — Fall 2023
Python re-implementation of the Torch API — [GitHub]

- Achieved a **Python** re-implementation of Torch API, achieving 100% compatibility with native PyTorch code
- Implemented auto-differentiation by architecting a Tensor class for mathematical operation **overloads** in Python, formulating the **computational graph** to compute the derivatives, and efficiently employing backpropagation via **Topological Sort**
- Optimized tensor operations on CPUs, achieving a **70%** speed boost by harnessing **Numba JIT** parallelization
- Committed to software engineering practices, including code styling with **black** and **flake8**, **pytest**-driven comprehensive unit testing

GenieChat, (React.js, Socket.io, Tailwind CSS, Express.js) — Summer 2023
Full-stack WhatsApp clone with ChatGPT API Integration — [GitHub]

- Developed a real-time chat application using **React.js** for the frontend and **Node.js** for the backend, utilizing **socket.io** for real-time bi-directional communication between the server and the client, and **PostgreSQL** for the database management system
- Enhanced loading performance by implementing **lazy loading**, resulting in a **65%** reduction in file bundle sizes
- Engineered secure and user-friendly login functionality with **Firebase**, enabling authentication through Google and GitHub
- Incorporated ChatGPT via the **OpenAI API**, offering automated responses and real-time language translation, enhancing user interactions

MF-Net, (PyTorch) — Summer 2022
GAN-based generative deep learning model for stylized font design — [GitHub] [Paper]

- Designed and trained an end-to-end GAN-based model with **PyTorch** for generating font images in arbitrary styles from reference images
- Integrated a multi-level **attention** module to capture both local and global style features, further complemented by an adaptive **skip connection** to adjust the preservation of the content structure, resulting in a **19%** SSIM improvement, as confirmed through an **ablation study**
- Conducted rigorous benchmarking of the model against SOTA baselines on 858 stylized fonts, demonstrating a **17.3%** SSIM improvement in image distance, a **27.4%** mFID improvement in feature distance, and a **16.7%** improvement in user evaluation, compared to the SOTA model
- Published a paper as the **first author** on this project on ***ACM Multimedia 2022***, which has received **2 citations**