Zhihong (Cody) Jiang

New York, NY
cody.jiang@nyu.edu | +1 (917) 519-4588
GitHub | Website | LinkedIn

EDUCATION

New York University — **New York, NY**
M.S. in Computer Science (Merit Scholarship Recipient) — Expected Dec 2026
B.S. in Computer Science, Minor in Applied Mathematics, Minor in Game Design — Sep 2021 – May 2024

Elite Preparatory Academy — **Somerset, NJ**
Honor Class — Sep 2018 – May 2020

TECHNICAL SKILLS

Programming: Python, SQL, JavaScript, C++, C#, Golang
Data Science & ML: Pandas, NumPy, PySpark, Dask, XGBoost, PyTorch, TensorFlow, Transformers, SpaCy
Statistics & Modeling: Regression, Feature Engineering, ETL, Model Evaluation
Visualization: D3.js, Vega-Lite, Matplotlib, Streamlit
Data Systems & Tools: MySQL, MongoDB, Git, Docker, Linux, REST APIs

DATA SCIENCE / ANALYTICS EXPERIENCE

Steam Review Radar — Platform Behavior & Signal Quality — NYU Big Data Project — **New York, NY**
Data Science / Analytics Contributor — Sep 2025 – Dec 2025

- Built large-scale review analysis workflows to study review helpfulness, platform behavior, and manipulation patterns in Steam review data.
- Implemented feature-generation and data processing pipelines for helpfulness prediction using **PySpark** and downstream modeling.
- Compared statistical and ML approaches (including **Negative Binomial regression** and **XGBoost**) for robustness and interpretability of helpfulness signals.
- Developed spam/duplication detection and reviewer archetype analysis pipelines to improve signal quality assessment.
- Designed evaluation metrics to reduce exposure bias and support more reliable platform-level inference beyond raw prediction accuracy.

AI-Powered Interview Simulation Tool — **New York, NY**
Applied Data / ML Engineering Assistant — Oct 2023 – Mar 2024

- Built and integrated NLP/audio analysis components to evaluate response quality and generate structured feedback for interview practice users.
- Implemented Python-based scoring workflows using **PyTorch, Transformers, Torchaudio, and SpaCy**.
- Expanded analysis coverage by incorporating speech/audio structure signals and nonverbal performance indicators into the evaluation pipeline.
- Improved usability and adoption through user-facing features and documentation for result interpretation.

Anhui Yuntai Transportation Development Limited — **Anhui, China**
Software Engineering Intern (Data & Operations Analytics Focus) — Jun 2023 – Aug 2023

- Built a C++ data management module for route, revenue, trip duration, and passenger-flow records to support operational reporting.
- Supported route and operations analysis for dispatch / car-rental workflows and optimization recommendations.
- Contributed to process improvements associated with an estimated **~15% efficiency gain** in operations workflow.

SELECTED DATA / ANALYTICS PROJECTS

AI-Generated Code Technical Debt Analysis System — **New York, NY**
Python, Dask, Git, Docker, Static Analysis, Visualization — Sep 2025 – Present

- Designed an incremental, diff-based analytics system for multi-branch repositories to quantify structural anomalies and AI-generated code risk patterns over time.

- Built lineage-aware data processing across commits/branches, enabling repeatable analysis on large code histories instead of one-off scans.
- Implemented fault-tolerant caching and commit-delta execution that reduced repeated analysis turnaround time from hours to minutes.
- Developed detectors for duplication bursts, dependency drift, unreachable code, and API inconsistencies; aggregated signals into repository-level risk views.
- Created dashboards for hotspot analysis, diff timelines, and trend monitoring to support engineering prioritization.

NVIDIA Options Market Pipeline & Visual Analytics (NYU Information Visualization Project) **New York, NY**
Python, Pandas, D3.js, Vega-Lite, Data Cleaning, Feature Engineering Sep 2025 – Dec 2025

- Built an end-to-end analytics pipeline for NVIDIA options-chain data (2020–2022), transforming raw option records into analysis-ready datasets for risk and market behavior analysis.
- Processed **2M+ option-chain rows** with ETL workflows including normalization, schema alignment, strike-distance features, and moneyness / DTE bucketing.
- Implemented calls/puts data integration and type standardization to improve consistency across contracts and reduce downstream analysis errors.
- Developed interactive visual analytics (D3/Vega-Lite) for risk-return exploration, temporal pattern comparison, and anomaly inspection across contract cohorts.
- Identified unstable behavior concentrated in near-expiry contracts and extreme strike regions, enabling targeted investigation of volatility-driven edge cases.
- Designed the project for reproducible analysis workflows so feature transformations and visual views could be iterated consistently across multiple hypotheses.

Self-Supervised World Model (JEPA) for Sequential Prediction (NYU Deep Learning Project) **New York, NY**
PyTorch, Self-Supervised Learning, Sequential Modeling Sep 2024 – Dec 2024

- Implemented a recurrent JEPA-style self-supervised model for future latent-state prediction on sequential visual/action inputs.
- Built training/evaluation workflows for preprocessing, experiment iteration, and cross-configuration comparison.
- Tested regularization strategies (VicReg/BYOL-inspired constraints) to improve representation stability and generalization.
- Designed probe-style evaluations (e.g., collision/unseen layout scenarios) to analyze model behavior beyond in-distribution trajectories.

ADDITIONAL DS/ANALYTICS PROJECTS

Sentiment & Toxicity Classification Tool *(BERT, Python, Streamlit)*
- Built an NLP app for sentiment/toxicity inference with confidence-score inspection for quick output review.
- Integrated preprocessing, inference, and visualization into a lightweight analysis workflow.

Anonymous Course Reviews Platform *(MongoDB, Express, React, Node.js)*
- Built a full-stack platform for structured storage and retrieval of course review data.
- Designed schemas and query workflows for filtering/search, supporting rating/text analytics use cases.

Airline Booking System *(SQL, Flask, Relational Databases)*
- Built a relational booking system with normalized schemas for users, flights, and reservations.
- Implemented SQL queries for search/booking workflows using joins, constraints, and validation logic.

Task & Reminder Backend Services *(Node.js, JWT, REST APIs, CRON)*
- Built backend services with authentication, CRUD APIs, and scheduled job execution.
- Implemented data models and recurring reminder logic for event-driven workflows.

SELECTED COURSEWORK

Data Analysis · Big Data · Information Visualization · Machine Learning · Deep Learning · Algorithmic Machine Learning and Data Science · Databases · Software Engineering · Design & Analysis of Algorithms · Human Computer Interaction