ZICHAO(ALEX) ZHANG

New York, NY | (332) 201-6618 | zz3491@cims.nyu.edu | zichaozhang.com

SUMMARY

Experienced data scientist specializing in computer vision, NLP/LLMs, and tabular ML, pairing end-to-end modeling with elegant data visualization and stakeholder-ready slides. Focus on reproducible Python/SQL pipelines and interpretable, robust modeling.

EDUCATION

New York University, Center for Data Science, New York, NY *Sep 2025 – May 2027 (Expected)*

Master of Science in Data Science

New York University, College of Arts and Science, New York, NY *Sep 2021 – May 2025*

Bachelor of Arts: Joint Major in Data Science and Computer Science, minor in Business Studies; GPA: 3.75/4.00

- Honors: Dean's List (all years); Founder's Day Award

SKILLS and CERTIFICATIONS

- Programming/Analytics: Python, SQL, Java, C, JavaScript/TypeScript, Assembly; Pandas, NumPy, Jupyter, MongoDB, Git, Tableau.
- Machine Learning/AI: scikit-learn, PyTorch, TensorFlow, XGBoost, Hugging Face Transformers, CNN, RNN, LLM, RAG
- Tools/Cloud: Matplotlib, Seaborn, FastAPI, React, Node.js, JSON, B CUDA (basic), AWS, NYU HPC.
- Certifications: Deep Learning Specialization; Machine Learning Specialization; Google Data Analytics Professional Certificate.

EXPERIENCE

Teaching Assistant (Causal Inference, DS-UA 201), NYU Center for Data Science, New York, NY *Sep 2025 – Present*

- Led weekly recitations and office hours for ~30 students, developing and delivering lecture slides and Jupyter labs.
- Authored problem sets and grading rubrics to improve consistency of instruction and assessment.

Data Analyst Intern, Qihoo 360, Beijing, China *Jun 2025 – Aug 2025*

- Collaborated with faculty to design an end-to-end AI/ML course pipeline (curriculum, datasets, capstone project, auto-grading), reducing instructor setup time by ~30% and enabling 100% grading uptime during pilot.
- Benchmarked and optimized a GPU-backed Jupyter/LLM environment (scikit-learn, XGBoost), creating pre-launch checklists and resource guidelines to ensure stable performance.

Research Assistant, CN3 Lab, NYU Langone, New York, NY *May 2024 – May 2025*

- Adapted Neural Data Transformer-2 (NDT-2) and Pi-VAE to decode ~330 CA1/VISp neurons to movie frames, reducing MAE by 12% (CA1) / 23% (VISp) vs. spline tuning curves.
- Pre-processed and aligned 54k spike events with 1,800 s video stimuli; implemented Bayesian Monte Carlo decoder and 10-fold CV to confirm robustness.

Undergraduate Researcher, NYU Center for Data Science, New York, NY *Sep 2024 – May 2025*

- Built a unified Python library extending TM-Vec for protein homology, streamlining data pipelines, training, and benchmarking across 380 K sequences; fine-tuned 1.2 B-parameter protein LLMs on NYU Greene HPC, reducing MAE by ~9× (to 0.003).
- Visualized 2048-D embeddings with PCA and t-SNE, revealing distinct alpha, beta, alpha/beta clusters.

PROJECTS

Learning from 100 Images: Data-Efficient Computer Vision with Explainability *Mar 2025 – May 2025*

- Achieved 98% accuracy on MNIST using only 100 training images by applying data condensation and augmentation to train a compact CNN in PyTorch.
- Evaluated model interpretability with Grad-CAM and Integrated Gradients, observing moderate feature alignment (top-saliency IoU ≈ 0.33) versus a full-data model; developed a reproducible CLI tool with visual summaries.

Auditing Fairness in Home Credit's Automated Decision System (ADS) *Mar 2024 – May 2024*

- Audited a loan-approval model (120+ features), uncovering a 20-percentage-point accuracy gap favoring male applicants.
- Designed an imputation and feature-pruning pipeline that improved AUC-ROC from 0.68 to 0.90 (+32%) while mitigating bias.

Deepfake Facial Imagery Detection *Feb 2024 – May 2024*

- Led a 3-member team to develop a mobile-optimized deepfake detector (MobileNetV1) with 94% test accuracy on a 140K-image GAN dataset.
- Enhanced cross-dataset generalization via style-diverse augmentation and error-level analysis, raising detection rates from 52% to 100% on Stable Diffusion images and to 84% on an unseen GAN dataset.
- Used Grad-CAM for model interpretability and presented findings to ~50 peers and faculty; project was ranked top of class.

COMPETITION

- RAISE 2024 (Rutgers University) – 3rd place among 100+ teams; NLP sentiment analysis on 10k+ AI-related headlines.