Yucheng (Cecilia) Zhang

737-343-4836 | ceciliayczhang@gmail.com | portfolio | linkedin/yczcecilia | github/yczhangcc | Google Scholar

Education

Northwestern University Aug. 2026 – May 2028
Master of Science in Computer Science Evanston, IL

The University of Texas at Austin Aug. 2023 – May 2026
Bachelor of Science in Mathematics, Minor in Computer Science | GPA: 3.74/4 Austin, TX
Relevant Courses: Data Structure, Stochastic Process, Elements of Software Design, Data Analytics, Principles of Database System, Discrete Math, Real Analysis, Probability and Statistic, Artificial Intelligence, Elements of Networking

Research Experience / Projects

Graduate Research Assistant — SPICE Lab

Northwestern University | Advisor: Prof. Karan Ahuja | Tools: Python, PyTorch, Slurm Aug. 2026 – Present

- Built a unified pipeline converting ∼1,490 h of human motion from 8 datasets (optical mocap, Xsens, multi-view video) into SMPL and simulated 6-IMU signals, with person-level train/test splits and per-dataset validation to prevent leakage.
- Designed a 60-run data-scaling study of the DynaIP pose model (5 data sizes × 4 dataset groups × 3 seeds); combined-dataset training cut simulated error to 3.5 cm MPJPE vs. 6.2–7.2 cm for the best single-dataset models.
- Identified the sim-to-real gap as the main bottleneck: real-IMU error stayed at 10–14 cm regardless of data size, traced to high-frequency sensor energy missing from simulation (8–40% vs. 0.4–3% above 5 Hz).
- Second author on an ACL ARR submission showing that SyncNet confidence (LSE-C), the standard lip-sync metric, does not predict lip readability; benchmarked 6 generators on English and Mandarin translated speech.

CodeAssist Project

University of Texas at Austin | Mentor: Prof. Kia Teymourian | Tools: Flask, Docker Aug. 2025 – Oct. 2026

- Developed modular backend architecture to integrate multiple AI providers (OpenAI, Hugging Face) via API routing, reducing vendor dependency and optimizing performance–cost tradeoffs.
- Implemented dynamic model-switching logic for large-scale deployments, improving scalability, reliability, and adaptability across multi-cloud environments.
- Debugged and enhanced front-end components in React, fixing feature update issues such as the "updated list" and "course description" display to ensure seamless UI–backend integration.
- Collaborated on system refactoring and Docker containerization to improve maintainability, reproducibility, and cross-team development efficiency.

Researcher & TA — Biological Data Analytics

University of Texas at Austin | Mentor: Prof. Becca Young | Tools: R, Linux Jan. 2024 – Dec. 2024

- Processed and analyzed 250GB of RNA-seq data using Linux, Docker, and AWS Cloud, optimizing computational workflows to improve processing speed by 6% and reduce storage costs by 5%.
- Applied bioinformatics pipelines (Salmon, RSEM, Trinity, samtools) for transcript quantification, achieving a 98.7% mapping rate and ensuring reproducibility across multiple datasets.
- Performed differential expression analysis with Voom + Limma, identifying 48 key genes across time-series samples and applying permutation-based FDR to enhance statistical robustness.
- Automated R-based workflows for normalization, PCA visualization, and volcano plot generation using ggplot2, improving interpretation efficiency by 15%.
- Collaborated with peers as a teaching assistant to train incoming FRI students on RNA-seq pipelines, promoting reproducibility and data literacy in early-stage biological research.

Research Member — Computer Vision & ML (Kiwifruit Quality Grading)

Huazhong Agricultural University | Mentor: Prof. Yan Tong | Tools: Python, scikit-learn Aug. 2022 – May 2025

- Curated and engineered a 53-dimensional multi-feature dataset (12 color, 16 texture, 12 shape, 13 defect dimensions) from 1,701 labeled kiwifruit images across 3 quality classes.
- Validated feature independence via Pearson correlation (<0.5) and mutual information analysis, and ranked feature contributions with permutation importance.
- Benchmarked 6 classical ML classifiers (SVM, Random Forest, Decision Tree, Logistic Regression, Naive Bayes, KNN) with 5-fold cross-validation; the best model (Logistic Regression) achieved 85.7% macro-F1.
- Supported validation and visualization of ResTNet (a hybrid ResNet–Vision Transformer), which achieved 99.5% macro-F1; published in *Expert Systems with Applications*.

Machine Learning Research Assistant — Neuroengineering Lab

Remote | Mentor: Dr. Kunqiang Qing | Tools: Python, MATLAB, MNE Jan. 2023 – May 2024

- Preprocessed 35 GB of clinical EEG/fNIRS recordings from 32+ subjects, removing artifacts with independent component analysis (ICA) and extracting power spectral density (via FFT) and functional connectivity features.
- Co-authored a peer-reviewed study comparing healthy controls with mild, moderate, and severe post-stroke cognitive impairment, using one-way ANOVA with permutation testing and Bonferroni correction to find significant biomarker differences ($p < 0.001$); published in *Frontiers in Neurology*.
- First-authored a systematic review of EEG and fNIRS in the digital assessment and therapy of Alzheimer's disease, published in *Frontiers in Neuroscience*.
- Built an EEG emotion-recognition pipeline and benchmarked RF, SVM, KNN, and XGBoost classifiers with cross-validation; published as a first-author paper at IEEE ICIPCA.

Work Experience

Machine Learning & Risk Analytics Intern May 2025 – Dec. 2025

Toyota Industries Commercial Finance, Inc. | Python, SQL Server, Azure, Power BI Dallas, TX

- Investigated rising truck-loan defaults by combining internal loan data with PayNet credit-bureau data (SQL Server ETL, feature engineering on credit inquiries, loan-to-value, and account tenure).
- Trained Logistic Regression and Random Forest default models; the combined data outperformed the internal-only baseline, reaching AUC >0.82 across Class 8 truck, trailer, and used-vehicle segments.
- Backtested existing risk tiers against actual defaults and found the lowest tiers underestimated risk (up to 3.7× higher default rates); redrew tier boundaries, tightening predicted-vs-actual alignment.
- Recommended 2 decision-time features for credit approval and automated monthly model refreshes with Power BI dashboards for stakeholders.

Financial Data Development Engineer – Fraud Detection Jul. 2024 – Aug. 2024

Caitong Securities Ltd. | SQL, Oracle, Stored Procedures Hangzhou, China

- Built SQL ETL pipelines over 10M+ daily transactions, engineering behavioral features (transfer timing, test-then-transfer patterns, account dormancy) with window functions and stored procedures.
- Designed a weighted risk-scoring model across 10+ behavioral indicators to flag fraud and money laundering.
- Tuned weights and thresholds by backtesting on known fraud cases; A/B tests vs. the existing model cut false positives 2%.

Data Science & AI Engineering Intern – Customer Analytics Nov. 2023 – Jun. 2024

Joblogic-X Corporation | Python, Prophet, XGBoost, CatBoost, Optuna, Docker, MongoDB Dallas, TX

- Forecasted daily demand for 19+ dishes with a Prophet + XGBoost pipeline (13 engineered features, Optuna tuning) on 2+ years of sales data, reaching R^2 up to 0.98 and MAPE as low as 7.4% on held-out data.
- Scored social media influencers by scraping 500+ posts into a weighted KOC model, finding that image posts drove 5.6× more engagement than video, which guided marketing targeting.

Data Scientist Assistant Intern – Image Restoration Jul. 2023 – Sep. 2023

Software Engineering Center, Chinese Academy of Sciences Beijing, China

- Deployed and optimized image restoration modules (VAE loss, Gaussian filters, LSGAN) on cloud servers via SSH, improving workflow stability and scalability for large-scale image reconstruction.
- Integrated GFPGAN API for automated face restoration and color correction, and containerized the system with Git for reproducible, cross-platform deployment and performance benchmarking.

Leadership & Community Involvement

Leader — Austin Chinese Campus Christian Fellowship Aug. 2023 – Jun. 2026

University of Texas at Austin

- Organized weekly Bible study sessions, retreats, and cultural activities such as hiking trips and traditional Chinese festival celebrations, fostering spiritual growth and cultural connection among members.
- Coordinated logistics for transportation and event scheduling, improving group communication and ensuring inclusivity for members with diverse backgrounds and schedules.

Volunteer — Art From the Streets Jan. 2026 – Aug. 2026

Austin, TX

- Supported community-based art programming for artists experiencing homelessness or housing insecurity, helping create a welcoming and respectful environment for participants.
- Assisted with event preparation, artwork organization, and visitor engagement, contributing to accessible public art experiences and community outreach.

Publications

1 Jin, Z., **Zhang, Y.**, Islam, R., & Ahuja, K. (2026). Synchrony is not readability: Evaluating lip-sync metrics on translated visual speech in English and Mandarin. *ACL ARR* (under review). [OpenReview]

2 Miao, H., Wang, Z., **Zhang, Y.**, … & Tong, Y. (2026). New kiwifruit quality grading methods based on image multi-feature fusion and ResTNet model. *Expert Systems with Applications*, 297, 129507. (IF = 7.5) [DOI]

3 Xu, M., **Zhang, Y.**, Zhang, Y., Liu, X., & Qing, K. (2024). EEG biomarkers analysis in different cognitive impairment after stroke: An exploration study. *Frontiers in Neurology*, 15, 1358167. (IF = 2.8) [Link]

4 **Zhang, Y.**, Zhang, Y., Jiang, Z., Xu, M., & Qing, K. (2023). The effect of EEG and fNIRS in the digital assessment and digital therapy of Alzheimer's disease: A systematic review. *Frontiers in Neuroscience*, 17, 1269359. (IF = 3.2) [Link]

5 **Zhang, Y.** (2023). EEG-Based on Emotion Recognition Using Machine Learning. In *IEEE ICIPCA* (pp. 226–232). [IEEE]

6 Li, D., Zhang, Z., & **Zhang, Y.** (2023). Public compassion under the mediation of suffering: Taking the videos of "Second Uncle" on social media as an example. *Media and Communication Research*, 4(5), 46–58. [Link]