

# Yichi Zhang

1468 Stonybrook Rd – Stony Brook, NY 11790 – USA  
☎ +1 934 255 9272 • ✉ yichi.zhang@stonybrook.edu  
🌐 yichixiaoju.github.io/YichiZhang.github.io

## Education

Stony Brook University (SBU)

B.Sc. Applied Math and Statistics, GPA: 3.96/4.00

College of Engineering and Applied Sciences

08/2024 – Present

Anhui University

B.Sc. Applied Statistics (Expected 2026), WPGA: 3.77/5.00 (87.89/100)

Stony Brook Institute at Anhui University

09/2022 – 06/2024

Honor: Academic Excellence Scholarship (Top 10%, 2023)

Skills: Python (PyTorch, NumPy, Pandas, HuggingFace, RDKit), Java, MATLAB, R,  $\LaTeX$ , Markdown

## Publication

ACL 2025: X. Liu, Y. Ding, J. Qu, **Y. Zhang**, W. Gao, Y. Liu. *RL-Guider: Leveraging Historical Decisions and Feedback for Drug Editing with Large Language Models*. The 63rd Annual Meeting of ACL. Accepted.

## Research Experience

Advisor: Prof. Yi Liu

SBU

Improved Molecule Generation Using LLMs in Genetic Algorithms

01/2025 – 05/2025

- Conducted **MolLEO** experiments (**Gemini 1.5 Pro**, **ChemDFM**) across 12 molecular optimization tasks.
- Designed evaluation using **Top-10 AUC** against baselines (**GPT-4o**, **Graph-GA**) with **RDKit** validation.
- Analyzed comparative performance: Gemini 1.5 Pro excelled across broad tasks, ChemDFM effective on niche objectives.
- Concluded that **LLM-enhanced GA** outperformed heuristic methods in 9/12 tasks, highlighting scalability and adaptability.

Advisor: Prof. Yi Liu

SBU

RL-Guider: Drug Editing with LLMs

12/2024 – 01/2025

- Implemented a **reinforcement learning (RL) agent** interacting with **LLMs** to guide molecular editing decisions.
- Applied **NumPy/Pandas** for data preprocessing and **RDKit** for chemical visualization.
- Designed **interpolation-based fitting** to analyze experimental results.
- Achieved **SOTA** on **12/16 tasks** with LLAMA and **10/16 tasks** with Deepseek.
- Structured workflow using mind-mapping and case studies to validate robustness of RL-LLM integration.

Advisor: Prof. Qingpei Hu

Chinese Academy of Sciences

Robust Regression under Adversarial Corruptions

06/2025 – 08/2025

- Improved **CRR algorithm** by introducing **quantile-based thresholding** with linear/exponential decay strategies.
- Implemented baselines (**TRIP**, **TORRENT**, **CRR**) in **NumPy** and migrated to **PyTorch** for **GPU acceleration**.
- Designed synthetic datasets with varying contamination ratios; plotted **convergence/error curves** to benchmark results.
- Demonstrated accelerated convergence under adversarial noise, linking empirical findings to **SSC/SSS theoretical conditions**.
- Integrated coding practice with mathematical reasoning to ensure **theory-experiment consistency**.

Advisor: Prof. Weisheng Niu

Anhui University

Homogenization Theory for First-Order Equations

12/2023 – 04/2024

- Reviewed mathematical literature to reconstruct models for **periodic homogenization**.
- Applied **asymptotic expansion**, **energy estimates**, and **corrector construction** to derive homogenized problems.
- Estimated convergence rates and analyzed **solution stability**.
- Authored a full  $\LaTeX$  manuscript including model formulation, rigorous proofs, and conclusions.
- Gained strong foundation in **PDE theory** and homogenization techniques with formal academic writing.

## Professional Experience

UU-Paotui, Efficiency Innovation Dept.

AI Agent Intern

Zhengzhou, China

06/2025 – 08/2025

- Built and optimized automation workflows on **Coze** by integrating **LLMs (GPT-4o, Gemini, Sora)** with NLP prompt engineering.
- Developed automated product management tools for requirement identification and document analysis.
- Designed **text-to-PPT generation** system with intelligent layout optimization for reporting and marketing.
- Implemented **auto-theme video generation** tools by orchestrating APIs (**Doc2Video, Sora**), enabling multimodal content production.
- Improved workflow stability and backend robustness by debugging integration pipelines and optimizing scripts.

## Forge

Online

### BCG Data Science Job Simulation

09/2024

- Performed customer churn analysis using **Pandas/NumPy**.
- Conducted data visualization with **Matplotlib** to identify behavioral trends.
- Trained and tuned a **Random Forest classifier**, achieving **85% accuracy** in churn prediction.
- Delivered an executive-style summary report with actionable insights for business strategy.

## New Development KAILIN Ltd.

Zhengzhou, China

### Business Analytics Intern

07/2024 – 08/2024

- Conducted competitor pricing and promotion research; performed **regression** and **elasticity analysis** achieving **92% predictive accuracy**.
- Designed questionnaires and built target user profiles to refine product positioning.
- Automated data cleaning and preprocessing workflows with **Pandas**, boosting efficiency by **40%**.
- Visualized pricing correlations using **Matplotlib** (line charts, bar graphs, heatmaps) and created strategic mind maps.
- Synthesized industry-wide patterns into data-driven recommendations for revenue optimization.

## Independent Projects

---

### 05/2025: Generative Model with DDPM (Denoising Diffusion Probabilistic Model)

- Implemented a PyTorch-based **DDPM** pipeline to generate MNIST digits.
- Designed **forward diffusion** with linear  $\beta$  schedule and reverse sampling with a **U-Net**.
- Integrated **sinusoidal timestep encoding**, **EMA**, and dynamic progress tracking (tqdm).
- Visualized denoising trajectories at different time steps; conducted ablation on **learning rate**, **noise schedule**, and **batch size**.
- Achieved clear image synthesis showing effective **noise removal**.

### 07/2025: Implementing LLaMA2 from Scratch (Happy-LLM Tutorial Extension)

- Reproduced **LLaMA2 architecture** (multi-head attention, rotary embeddings, causal masking) in PyTorch.
- Built a custom **tokenizer** and positional encoding for sequence preprocessing.
- Trained a small-scale model on toy datasets; monitored **loss curves** and **perplexity**.
- Extended tutorial with **gradient clipping**, **learning rate scheduling**, and **mixed-precision training** to improve GPU stability.
- Validated generation quality through **prompt-based evaluation**.

## Contest

---

### Interdisciplinary Contest in Modeling (ICM)

#### Honorable Mention Award

01/2024

- Problem F: **Reducing Illegal Wildlife Trade**.
- Applied **Analytic Hierarchy Process (AHP)** and **TOPSIS** to rank potential partners for intervention strategies.
- Used **Principal Component Analysis (PCA)** to reduce redundancy and correlation analysis to identify key variables.
- Built a **multiple linear regression model** to quantify policy effectiveness pre- and post-implementation.
- Conducted **sensitivity analysis** to ensure robustness of results under parameter changes.

## Leadership & Activity

---

### Stony Brook Institute at Anhui Univ.

#### Community Service Volunteer

10/2023 – 06/2024

- Participated in **public service activities** (clean-up, community support).
- Delivered **academic-interest lectures** for children.

### Applied Statistics Class 2

Anhui University

#### Vice Monitor

09/2022 – 06/2024

- Coordinated class communication and **academic activities**.
- Supported faculty-student interaction and organizational tasks.