

Natalie Cao

735 Campus Drive, Stanford, CA 94305 | (+86) 178-0132-9912 | nc323@stanford.edu

EDUCATION

Stanford University

Stanford, CA

Master of Science in *Management Science & Engineering*

Sep 2023 - May 2025

- GPA: **3.93/4.00**; Major GPA: **4.00/4.00**
- Vice President at Department of Career Development of ACSSS (Association of Chinese Students and Scholars)
- **Relevant Courses:** Natural Language Processing, Machine Learning, Causal Inference, Decision Analysis, Probabilistic Analysis, Optimization (advanced).

Duke Kunshan University & Duke University

Kunshan, China & Durham, NC

Bachelor of Science in *Applied Mathematics*

Aug 2018 - May 2022

- GPA: **3.85/4.00**; Major GPA: **3.90/4.00**
- Scholarship Recipient (**top 5%**) & Dean's List (awarded every semester)
- **Relevant Courses:** Ordinary Differential Equations (ODE), Partial Differential Equations (PDE), Dynamical Systems, Numerical Analysis, Stochastic Modeling.

PUBLICATIONS & MANUSCRIPTS

- “3D Reconstruction of the Spatial Transcriptomics with Neural Field.” Yifan Lu, **Natalie Cao**, and Xiaojie Qiu. To be submitted to *Nature BioTechnology*, in preparation.
- “Towards a Predictive Virtual Embryo with Genomics and AI.” **Natalie Cao**, Xiaojie Qiu. *Nature Methods*, under review.
- “Predictive Modeling of Single Cell Transcriptomic Dynamics with Dynamo.” Zehua Zeng, Sichao Yu, Ke Ni, Yan Zhang, Ukyeon Shin, Cinlong Huang, **Natalie Cao**, Jonathan Weissman, Jianhua Xing, Xiaojie Qiu. (2025). *Nature Protocols*, under review.
- “A Financial Decision Support Optimization System Based on Artificial Intelligence and Data Mining Algorithms.” **Natalie Cao**. IEEE EEBDA 2024. DOI: 10.1109/EEBDA60612.2024.10485835
- “The Application of AI Large Models.” Contributing Author. ISBN: 9787111738787.

RESEARCH EXPERIENCE

Virtual cell/embryo and single cell/spatial genomics dataset modeling — Qiu Lab

Stanford, CA

Advisor: Professor Xiaojie Qiu, Dept. of Genetics & Computer Science, Stanford University

Mar 2025 - Present

Methods and Tools: scRNA-seq, spatial transcriptomics, RNA velocity/kinetics, PyTorch, Biorender

- Co-developed a differentiable framework “neural_field”, a new differentiable framework for 3D embryo reconstruction using neural fields. Enhanced algorithmic stability and added Transformer-based positional encodings, improving slice alignment accuracy.
- Led the brainstorm, design, and writing of “Towards a Predictive Virtual Embryo with Genomics and AI”. Proposed a conceptual framework of virtual embryo that goes beyond virtual cell by a multi-scale, cross-modal predictive modeling of single cell and spatial genomics with hierarchical graph neural network, flow matching and transformers.
- Co-authored on the manuscript of “Predictive Modeling of Single Cell Transcriptomic Dynamics with Dynamo”. Contributed to developing a general framework for RNA velocity vector field learning, downstream differential geometry analyses and non-trivial prediction of optimal reprogramming prediction and in silico genetic perturbation predictions.

Deep Learning for Consumption and Production Management in the Circular Economy

Stanford, CA

Advisor: Professor Yu Ding, Stanford Graduate School of Business

Apr 2024 - Jul 2024

Methods and Tools: NLP, BERT, GPT, Deep Learning, Data Analysis, Python (TensorFlow/PyTorch)

- Developed BERT/GPT pipelines, techniques transferable to biomedical literature mining (PubMed triage, assay/phenotype tagging) and evidence synthesis.
- Applied LSTM and transfer learning; experimental rigor applicable to sequence/time-series biological signals and phenotype prediction.

Reinforcement Learning for Dynamic Systems — HPC-AI Lab

Singapore

Advisor: Professor: Professor Yang You, Dept. of Computing, National University of Singapore

Apr 2023 - Sep 2023

Methods and Tools: Reinforcement Learning, PPO, PPG, Project Management, AI Model Deployment, Python

- Developed the Variance-Regularized Value Optimization (VRVO) method, improving generalization stability over PPO/PPG in high-variance RL environments. Derived theoretical regularization under stochastic gradient dynamics; validated on the OpenAI Procgen benchmark.
- Co-authored paper “Improving Generalization of Reinforcement Learning via VRVO” (under review); contributed to The Application of AI Large Models (ISBN: 9787111738787).

PROFESSIONAL EXPERIENCE

Sahara AI (*decentralized AI blockchain platform*) **San Francisco, CA**
Machine Learning Engineer Summer Intern Jun 2024 - Sep 2024

- Built a real-time Twitter sentiment analysis pipeline with BERT and transformers, achieving 50% fine-grained emotion recognition accuracy.
- Optimized hyperparameters through A/B testing, reducing inference time by 20% without accuracy loss.

Ping An Technology (*fintech and AI solutions provider*) **Beijing, China**
AI Data Operation Intern (Part-time) Aug 2022 - Aug 2023

- Improved loan default prediction accuracy by 12% using SHAP-based feature selection in the FinRisk Classifier. Built an LSTM-attention model for early repayment prediction (AUC = 0.85).

Sequoia Capital China **Beijing, China**
Investment Research Intern Oct 2022 - Apr 2023

- Developed a reinforcement learning-based asset allocation model integrating 1992–2020 macroeconomic data, increasing simulated portfolio returns by 12%. Presented findings to senior partners, influencing strategy.

OPEN SOURCE PROJECTS

Neural Field: Generative Framework for 3D Embryo Reconstruction **Stanford, CA**
Contributor (Private GitHub Repository) Sep 2025 - Present

- Contributed to aristoteleo/neural_field, a framework for reconstructing 3D embryo structures from sequential 2D spatial transcriptomics slices.
- Integrated Transformer-based spatial encodings and VAE modules, improving alignment accuracy, loss smoothness, and robustness on Stereo-seq datasets.
- Analyzed large-scale human and mouse embryo datasets to reveal spatiotemporal gene regulation and signaling dynamics underlying organogenesis and congenital diseases.

Dynamo-Release: Dynamic Modeling for Single-Cell Trajectories **Stanford, CA**
Open-Source Contributor (472 ★ GitHub Repository) Aug 2025 - Present

- Contributed to aristoteleo/dynamo-release, an open-source framework for modeling single-cell dynamics and RNA velocity analysis. Enhanced model compatibility and performance with AnnData ≥ 0.10 , resolving key import and dependency issues for broader user adoption.
- Improved data preprocessing and visualization modules, optimizing trajectory inference and phase portrait generation for large-scale single-cell datasets.

SENTINEL: Sophisticated Ensemble for Tracing and Identifying Non-Authentic Language **Stanford, CA**
Custom Project for CS 224n: Natural Language Processing with Deep Learning Apr 2024 - June 2024

- Developed SENTINEL, a heterogeneous ensemble framework integrating deep learning models (BERT, RoBERTa) with classical algorithms (MNB, SGD, LightGBM, CatBoost) for AI-generated text detection, achieving 94.98% accuracy and 0.9685 AUC-ROC on the Kaggle dataset.
- Engineered a custom preprocessing pipeline and hybrid feature extraction method combining RoBERTa embeddings with TF-IDF n-grams, enhancing model performance across diverse text types and styles.

TEACHING EXPERIENCE

MS&E 152&252: Decision Analysis I & II **Stanford, CA**
Teaching Assistant Aug 2024 - Dec 2024, Jan 2025 - Mar 2025

MS&E 180: Organizations: Theory and Management **Stanford, CA**
Teaching Assistant June 2024 - Aug 2024

ADDITIONAL INFORMATION

Programming Languages: Python (TensorFlow, PyTorch), R, SPSS

Machine Learning & AI: Reinforcement Learning (PPO, PPG, LSTM), NLP, Transformers (BERT, GPT, DistilBERT), Predictive Modeling, Time Series Forecasting

Data Analysis & Visualization: Statistical Modeling, Feature Selection (SHAP values), A/B Testing