Zhengxu (Jason) Yan

Stanford, CA 94305 | (509) 330-6699 | jasonyan@stanford.edu | Citizenship: U.S. Citizen | zhengxuyan.github.io

EDUCATION

Stanford University — Expected Graduation: April 2027
Master of Science in Computer Science — GPA: 4.00/4.00

University of California, Berkeley — August 2021 - May 2025
Bachelor of Arts in Computer Science and Data Science — GPA: 3.95/4.00, High Distinction

PEER-REVIEWED PUBLICATIONS

• **VERINA: Benchmarking Verifiable Code Generation**
Ye, Z., **Yan, Z.**, He, J., Kasriel, T., Yang, K., & Song, D. *International Conference on Learning Representations (ICLR) 2026, April 23–27, 2026, Rio de Janeiro, Brazil.* | ICLR Proceedings | arXiv

• **Vero: Can AI Agents Build Formally Verified Software Repositories?**
Ye, Z.*, Lou, H.*, Sun, Y.*, Song, P., **Yan, Z.**, Kasriel, T., Zhang, Q., Yang, K., Kong, S., He, J., & Song, D. *Conference on Neural Information Processing Systems (NeurIPS) 2026, Evaluations and Datasets Track. Accepted as a poster.* | arXiv | Project

• **A Deep Reinforcement Learning Framework for Multi-Dimensional Adaptability in Building Operations**
Schaumann, D., **Yan, Z.**, Haiman, E., & Kalay, Y. E. *Journal of Building Engineering, 124, 115993 (2026).* DOI

• **Understanding Older People's Voice Interactions with Smart Voice Assistants: A New Modified Rule-Based Natural Language Processing Model with Human Input**
Yan, Z., Dube, V., Heselton, J., Johnson, K., Yan, C., Jones, V., Blaskewicz Boron, J., & Shade, M. *Frontiers in Digital Health, 6, 1329910 (2024).* DOI

• **Reducing Loneliness and Improving Social Support among Older Adults through Different Modalities of Personal Voice Assistants**
Jones, V. K., Yan, C., Shade, M. Y., Boron, J. B., **Yan, Z.**, Heselton, H. J., Johnson, K., & Dube, V. *Geriatrics, 9(2), 22 (2024).* DOI

PREPRINTS & MANUSCRIPTS

• **What Your Policy Cannot See: Robot Data Curation via Observability Estimation**
Gao, T.*, **Yan, Z.***, Mahajan, A., Bhowmik, R. T., Hejna, J., Sadigh, D., & Finn, C. *Under review at IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 2027. IRIS Lab, Stanford University.*

• **NoveltyRank: A Retrieval-Augmented Framework for Conceptual Novelty Estimation in AI Research**
Yan, Z., Li, H., & Feng, Y. *Preprint. Stanford University.* | arXiv

• **DiT-Serve: An Efficient Serving Engine for Diffusion Transformers**
Luo, M., Hao, A., **Yan, Z.**, Cao, C., & Nguyen, Q. L. N. *Manuscript. Sky Computing Lab, UC Berkeley.*

• **Auditing Added Information in NLP Dataset Papers with Dataset Contribution Units**
Yan, Z., Fang, J., Spangher, A., Baumann, J., & Yang, D. *Manuscript in preparation. SALT Lab, Stanford University.*

RESEARCH EXPERIENCE

Graduate Research Assistant (Supervised by Dr. Chelsea Finn) — November 2025 - Present
Intelligence through Robotic Interaction at Scale Lab (IRIS), Stanford AI Lab (SAIL) — *Stanford, CA*

Information-Theoretic Data Curation for Vision-Language-Action Models
• Co-led the design and development of observability-aware data curation methods for vision-language-action models, from research formulation and method design to implementation and evaluation.
• Co-designed the scoring framework and trajectory-level aggregation methods to identify demonstrations with insufficient task-relevant visual information.
• Implemented scoring models and trained and evaluated policies on curated demonstration subsets against standard baselines.

Extensions to Feedback Descent
• Proposed and implemented variants of Feedback Descent that improved performance over the original method.

Graduate Research Assistant (Supervised by Dr. Diyi Yang and Dr. Alexander Spangher) — November 2025 - Present
Stanford Social and Language Technologies Lab (SALT), Stanford AI Lab (SAIL) — *Stanford, CA*
• Independently developed Dataset Contribution Units (DCUs), a typed representation of how a dataset extends prior work.
• Defined contribution dimensions spanning task, domain, annotation, scale, evaluation, and availability.
• Developed and evaluated type-routed retrieval and support-aware reranking to connect contribution claims with supporting or

conflicting prior evidence.
• Built the paper bank, extraction pipeline, retrieval benchmark, and longitudinal-analysis tools for auditing NLP dataset contributions and adoption.

Research Assistant (Supervised by Dr. Ion Stoica) February 2024 - October 2025
UC Berkeley Sky Computing Lab *Berkeley, CA*
• Led a five-person team developing DiT-Serve, an efficient serving system for diffusion transformers.
• Integrated Ring Attention and Brick Attention for distributed long-sequence video generation.
• Designed continuous batching and request scheduling for mixed-length, mixed-resolution workloads.
• Built shared model interfaces and multi-GPU benchmarks for memory, latency, and throughput.
• Contributed to DiT-Serve's system design, implementation, and experimental evaluation.

Research Assistant (Supervised by Dr. Yehuda Kalay and Dr. Davide Schaumann) February 2024 - September 2025
UC Berkeley College of Environmental Design *Berkeley, CA*
• Led a three-person team developing a multi-objective reinforcement-learning framework for adaptive building operations.
• Formulated cardiac-lab control as joint optimization over environmental and operational constraints.
• Built the simulation and implemented PPO and Q-learning controllers for multi-dimensional adaptation.
• Directed experiments and analysis for the resulting Journal of Building Engineering publication.

Machine Learning Engineer Lead (Supervised by Dr. Stefano Bertozzi) January 2024 - September 2025
UC Berkeley School of Public Health *Berkeley, CA*
• Led ML engineering for Rapid Reviews: Infectious Diseases, an automated preprint-triage system.
• Developed SciBERT- and LLM-based models for scope classification and conflict-of-interest screening.
• Designed human-in-the-loop workflows for domain experts to inspect and correct model outputs.
• Directed teams building ingestion, deduplication, and metadata normalization across preprint repositories.
• Collaborated with Berkeley Public Health and UCSF researchers on deployment in review workflows.

Research Assistant (Supervised by Dr. Dawn Song) 2024 - 2026
Berkeley RDI, UC Berkeley Department of Electrical Engineering and Computer Sciences *Berkeley, CA*
• Co-developed VERINA, an ICLR 2026 benchmark for machine-verifiable Lean code, specifications, and proofs.
• Built compiler-guided generation and repair pipelines with executable evaluation for each output type.
• Extended the work to Vero, the first repository-level benchmark for joint implementation and proof synthesis.
• Contributed benchmark curation, dependency-aware evaluation, and audits for flawed specifications or reference code.

Research Assistant (Supervised by Dr. Valerie Jones) April 2023 - September 2023
University of Nebraska-Lincoln *Lincoln, NE*
• Developed a human-in-the-loop NLP method for analyzing older adults' voice-assistant interactions.
• Designed annotation and error-review workflows that let domain researchers refine model rules.
• Contributed to empirical analysis and two peer-reviewed journal publications.

TEACHING & ACADEMIC SERVICE

Teaching Assistant for Large Language Model Agents (CS 194-196) (Supervised by Dr. Dawn Song) August 2024 - May 2025
UC Berkeley Department of Electrical Engineering and Computer Sciences *Berkeley, CA*
• Refined assignments and technical materials on tools, planning, memory, and multi-agent coordination.
• Reviewed student agent implementations and advised on system design, debugging, and evaluation.
• Helped organize the Berkeley RDI LLM Agents Hackathon and supported participating teams.

HONORS & SCHOLARSHIPS

High Distinction, University of California, Berkeley 2025
Sky Lab Summer Undergraduate Scholarship, UC Berkeley 2024
EECS Diversifying Access to Research in Engineering (DARE) Scholar, UC Berkeley 2024
Cheunkarndee Fund Summer Research Stipend, UC Berkeley URAP 2024

TECHNICAL SKILLS

Programming: Python, C/C++, Java, Go, Rust, SQL, JavaScript, Bash
Machine Learning: PyTorch, TensorFlow, Stable Baselines, OpenAI Gym, pandas, NumPy
LLM and Data: Hugging Face, AutoGen, retrieval and reranking pipelines, scientific document processing
Formal Methods: Lean, executable verification, compiler-guided proof repair
Systems: Docker, FastAPI, PostgreSQL, Git, AWS, Google Cloud