HAOSHENG (WOODY) GAN

Google Scholar: scholar.google.com/citations?user=26DHSpcAAAAJ
LinkedIn: linkedin.com/in/woody-gan
Email: woodygan@usc.edu

EDUCATION

University of Southern California *Aug 2022 - May 2026 (expected)*
B.S. in Computer Science and B.S. in Applied Mathematics
GPA: 4.0/4.0
Related courses: Mathematics of Machine Learning, Computer System, Theory of Computing, Internetworking, Statistics, Real Analysis, Combinatorics, Number Theory, Probability Theory

RESEARCH INTERESTS

I am interested in advancing **mechanistic interpretability** and developing **robust evaluation frameworks** of **multimodal models** incorporating **audio and vision**.

RESEARCH EXPERIENCE

Stanford SALT Lab (UGVI Program) *Supervisor: Prof. Diyi Yang*
Stanford, CA

CAVANA: Evaluating Large Audio Models by Human Preference *Jul 2025 - Present*

- Developed a model-agnostic voice agent scaffold for **Large Audio Models (LAMs)**, collecting $> 1,000$ human ratings from 10-minute goal-oriented conversations via Prolific.
- Conducted **meta-evaluation of LAM benchmarks** by correlating existing evaluations with human preference data to identify predictors of user satisfaction.
- Implemented **"anchor points" methodology** to create minimal evaluation sets, reducing developer overhead while maintaining predictiveness of user satisfaction.

AudioJudge: Understanding LAM Based Speech Evaluation [1] *Apr 2025 - Jul 2025*

- Conducted systematic study of **Large Audio Models as judges** across diverse speech evaluation tasks, demonstrating strengths and limitations for both example-level audio characteristics and system-level performance assessment.
- Developed frameworks achieving **0.91 Spearman correlation** with human preferences on speech-in-speech-out system evaluation.
- Performed comprehensive robustness analysis on acoustic noise resilience, verbosity bias, positional bias, and cross-modality consistency for LAM-based speech evaluation.

CAVA: Comprehensive Assessment for Voice Assistants [2] *Feb 2025 - Apr 2025*

- Contributed to benchmark curation assessing **Large Audio Model (LAM)** capabilities across dimensions critical for developing speech-in-speech-out voice assistants.
- Developed **turn-taking prediction task** using AMI meeting corpus to evaluate LAM ability to predict appropriate conversation entry points.

· Developed **speaker diarization task** using AMI corpus via Jaccard Error Rate with Hungarian algorithm-based matching.

SAEGull Group & USC Palms Lab
University of Southern California
Supervisors: Prof. Willie Neiswanger & Prof. Vatsal Sharan

Mechanistic Analysis of In-Context Learning Across Architectures *Jun 2025 - Present*

· Conducted systematic study of how different neural architectures (**LSTM**, **S4**, **Mamba**, **Transformers** with varied attention patterns) learn **linear regression problems** through in-context learning, analyzing layer-by-layer learning dynamics.

· Developing **theoretical framework** connecting architecture-specific inductive biases to optimization order, providing mathematical justification for observed in-context learning behaviors across model families.

Textual Steering Vectors for Multimodal LLM Enhancement [3] *Apr 2024 - Apr 2025*

· Extracted interpretable steering vectors from text-only LLM backbones to enhance visual reasoning capabilities in multimodal large language models.

· Achieved significant performance improvements: **+7.3%** spatial relationship accuracy and **+3.3%** counting accuracy on CV-Bench dataset.

· Demonstrated strong generalization to out-of-distribution datasets with up to **+15.8%** improvement on spatial tasks and **+34.2%** on counting tasks.

USC AIF4S *Supervisor: Prof. Mahdi Soltanolkotabi*
University of Southern California

ConceptMix++: Leveling Text-to-Image Benchmarking [4] *Oct 2024 - Apr 2025*

· Developed **gradient-based iterative prompt optimization** framework disentangling prompt phrasing from visual generation capabilities, achieving up to **20% performance improvement** across multiple diffusion models, revealing hidden model capabilities that standard benchmarks systematically underestimate.

· Demonstrated strong **cross-model transferability** of optimized prompts and identified category-specific performance patterns, with spatial relationships and shapes showing highest sensitivity to prompt formulation.

USC-Meta Center *Supervisor: Prof. Murali Annavaram*
University of Southern California

Privacy-Preserving LLMs with Token-wise Logit Mixing [5] *Apr 2024 - Nov 2024*

· Developed a novel **token-wise logit mixing** inference methodology for **privacy-preserving LLMs**, enabling **differential privacy** in **in-context learning** while preserving utility.

· Achieved state-of-the-art performance in privacy-preserving language tasks, including **99.3%**

summarization accuracy and **97.1% NLG accuracy** on **LLaMA-3.2 3B** while maintaining strong privacy bounds ($\epsilon = 4$).

Membership Inference Attacks on Private LLMs *Sep 2023 - Apr 2024*

- Implemented **Membership Inference Attacks (MIAs)** on differentially private LLMs trained on public datasets (DBpedia, SST-2, AGNews), systematically evaluating privacy vulnerabilities.
- Created a novel dataset by injecting datapoints simulating diverse categories of **sensitive information** into WikiText, enabling comprehensive analysis of privacy implications through **MIA** evaluations.

WORK EXPERIENCE

Shepherd Ventures *Apr 2024 - Oct 2024*
Financial Quant Intern & ML Team Co-lead *San Diego, CA*

- Engineered advanced ML models leveraging **K-Means Clustering**, **Random Forest**, **time series analysis**, and **Markowitz's mean variance optimization** to detect market regime shifts and predict tail events in **high-frequency data**.
- Developed adaptive **portfolio optimization** algorithms with real-time market signals, reducing downside risk by **60.7%** while maintaining benchmark-competitive returns.

Sports Media Inc *May 2024 - Aug 2024*
Software Engineering/AI Intern *Severance, CO*

- Deployed production-grade fine-tuned **LLMs** (**Gemma**, **LLaMA**) with **distributed microservices**, enabling seamless website integration serving **10,000+** users.
- Architected real-time **AI voice interaction** system integrating **Twilio Media Streams** with custom LLM endpoints for dynamic conversation processing.
- Designed and implemented **real-time content recommendation system** using **collaborative filtering** and **LLM embeddings**, improving user engagement by **40%**.

PUBLICATIONS / PREPRINTS / PROJECTS

[1] P. Manakul*, **Gan, Woody Haosheng***, M. J. Ryan, *et al.*, "Audiojudge: Understanding what works in large audio model based speech evaluation," *arXiv preprint arXiv:2507.12705*, 2025, * Equal contribution.

[2] W. Held, M. J. Ryan, A. Shrivastava, *et al.*, *Cava: Comprehensive assessment of voice assistants*, `https://github.com/SALT-NLP/CAVA`, A benchmark for evaluating large audio models (LAMs) capabilities across six domains: turn taking, instruction following, function calling, tone awareness, safety, and latency, 2025. [Online]. Available: `https://talkarena.org/cava`.

[3] **Gan, Woody Haosheng**, D. Fu, J. Asilis, *et al.*, "Textual steering vectors can improve visual understanding in multimodal large language models," *arXiv preprint arXiv:2505.14071*, 2025.

[4] **Gan, Haosheng**, B. Tinaz, M. S. Sepehri, Z. Fabian, and M. Soltanolkotabi, "Conceptmix++: Leveling the playing field in text-to-image benchmarking via iterative prompt optimization," *arXiv preprint arXiv:2507.03275*, 2025.

[5] J. Flemings, **Gan, Haosheng**, H. Li, M. Razaviyayn, and M. Annavaram, "Differentially private in-context learning via sampling few-shot mixed with zero-shot outputs," *arXiv preprint arXiv:2501.19287*, 2025.

ACHIEVEMENTS

Tau Beta Pi Membership, Engineering Honor Society — *Fall 2024*
CURVE fellowship, $2500, awarded by Viterbi School of Engineering, USC — *Fall 2024*
CURVE fellowship, $3000, awarded by Viterbi School of Engineering, USC — *Summer 2024*
CURVE fellowship, $1250, awarded by Viterbi School of Engineering, USC — *Spring 2024*
Phi Beta Kappa Membership, National Honor Society — *Spring 2024*
Academic Achievement Award, $4488, awarded by USC — *Spring 2024*
Academic Achievement Award, $4488, awarded by USC — *Fall 2023*
Dean's List, University of Southern California — *Fall 2022 - Fall 2024*

ACADEMIC APPOINTMENTS

Math 430: Theory of Numbers — *Fall 2024*
Teaching Assistant — *University of Southern California*

Math 126: Calculus II — *Spring 2024*
Course Grader — *University of Southern California*

Math 226: Calculus III — *Fall 2023*
Course Grader — *University of Southern California*