

# HAOSHENG (WOODY) GAN

Personal Website: <https://woodygan.github.io>

LinkedIn: [linkedin.com/in/woody-gan](https://www.linkedin.com/in/woody-gan)

Google Scholar: [scholar.google.com/citations?user=26DHSpcAAAAJ](https://scholar.google.com/citations?user=26DHSpcAAAAJ)

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, Internet-working, Statistics, Real Analysis, Combinatorics, Number Theory, Probability Theory*

## RESEARCH INTERESTS

---

**Natural Language Processing** and **Multimodal Models** incorporating audio and vision, with a focus on **mechanistic interpretability** and **robust evaluation frameworks**.

## RESEARCH EXPERIENCE

---

### Stanford SALT Lab (UGVI Program)

Supervisor: Prof. Diyi Yang

Stanford University, CA

*SODA: Scaling Open Discrete Audio Foundation Models [1]*

Oct 2025 - Feb 2026

- Pretrained **audio foundation models** jointly modeling interleaved audio and text tokens via next-token prediction, enabling a single backbone to support diverse audio and cross-modal capabilities.
- Conducted scaling law study for discrete audio models via **IsoFLOP analysis across 64 models** spanning  $3 \times 10^{18}$  to  $3 \times 10^{20}$  FLOPs, deriving optimal compute allocation and showing optimal data scales **1.6 $\times$  faster** than model size due to lower information density of audio tokens.
- Fine-tuned base models on downstream tasks including **voice-preserving speech-to-speech translation** and **ASR/TTS** to validate the flexibility of the SODA backbone across diverse audio tasks.

*HUMANS: Human-Validated Benchmark Subsets for LAMs [2]*

Jul 2025 - Jan 2026

- Developed and evaluated **10 subset selection methods** across 18 audio models and 40 tasks ( $\sim 16,000$  datapoints), showing **0.3% of items preserve 0.934 Pearson correlation** while dramatically reducing evaluation costs.
- Collected **776 human preference ratings** from 10-minute voice assistant interactions to validate benchmark subsets and conduct meta-evaluation of existing audio benchmarks.
- Trained **Ridge regression models** to predict human preferences from benchmark subsets, achieving **0.977 Pearson correlation** and providing practitioners an efficient proxy for human evaluation.

*AudioJudge: Understanding LAM Based Speech Evaluation [3]*

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 **0.912 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 [4]*

*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** and **speaker diarization tasks** using AMI meeting corpus, evaluating LAM ability to predict conversation entry points and identify speakers via Jaccard Error Rate.

**SAEGull Group & USC PALMS Lab**  
Los Angeles, CA

*Supervisors: Prof. Willie Neiswanger & Prof. Vatsal Sharan*

*Mechanistic Analysis of In-Context Learning Across Architectures*

*Aug 2024 - Present*

- Conducted systematic study of how different neural architectures (**LSTM, S4, RWKV, Mamba, Transformers** with varied attention patterns) learn **linear regression** through in-context learning, analyzing layer-by-layer learning dynamics.
- Proposed **hybrid architecture** from mechanistic insights, utilizing Transformer layers in early stages to develop task-specific representations and RWKV layers later for computational efficiency.
- 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 [5]*

*Apr 2024 - Apr 2025*

- Demonstrated **cross-modal transfer effect**: steering vectors from text-only LLM backbones effectively guide multimodal counterparts, enabling reuse of existing interpretability tools.
- Systematically evaluated three steering methods across MLLMs, achieving up to **+7.3% spatial accuracy** and **+3.3% counting accuracy** on CV-Bench with Mean Shift approach.
- Demonstrated strong out-of-distribution generalization with **+34.2% counting improvement** on CLEVR, outperforming LoRA fine-tuning while maintaining interpretability.

**USC AIF4S**  
Los Angeles, CA

*Supervisor: Prof. Mahdi Soltanolkotabi*

*ConceptMix++: Leveling Text-to-Image Benchmarking [6]*

*Oct 2024 - Apr 2025*

- Designed **gradient-inspired iterative prompt optimization framework** to disentangle prompt sensitivity from visual generation capabilities in text-to-image evaluation, revealing **rigid benchmarks underestimate performance by up to 20%**.
- Identified category-specific patterns where **spatial relationships and shapes benefit most from optimization**, revealing systematic underestimation of model capabilities in these visual domains.
- Demonstrated **cross-model transferability** of optimized prompts across DALL-E 3, Stable Diffusion 3.5, and Playground v2.5, indicating shared prompt preferences across diffusion architectures.

*Privacy-Preserving LLMs with Token-wise Logit Mixing [7]*

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 datasets by injecting datapoints simulating diverse categories of **sensitive information** into WikiText, enabling comprehensive analysis of privacy implications through **MIA** evaluations.

PUBLICATIONS / PREPRINTS / PROJECTS

---

- [1] P. Manakul, **Woody Haosheng Gan**, M. Bartelds, G. Sun, W. Held, and D. Yang, “Scaling open discrete audio foundation models with interleaved semantic, acoustic, and text tokens,” *arXiv preprint arXiv:2602.16687*, 2026.
- [2] **Woody Haosheng Gan**, W. Held, and D. Yang, “Putting HUMANS first: Efficient LAM evaluation with human preference alignment,” in *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026 Main)*, 2026.
- [3] P. Manakul\*, **Woody Haosheng Gan\***, M. J. Ryan, *et al.*, “Audiojudge: Understanding what works in large audio model based speech evaluation,” in *Proceedings of the 21st Conference of the European Chapter of the Association for Computational Linguistics (EACL 2026 Main)*, \* Equal contribution, 2026.
- [4] W. Held, M. J. Ryan, ..., **Woody Gan**, and D. Yang, *Cava: Comprehensive assessment of voice assistants*, <https://github.com/SALT-NLP/CAVA>, A benchmark for evaluating large audio models (LAMs) capabilities across six domains, 2025. [Online]. Available: <https://talkarena.org/cava>.
- [5] **Woody Haosheng Gan\***, D. Fu\*, J. Asilis\*, *et al.*, “Textual steering vectors can improve visual understanding in multimodal large language models,” in *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (ACL 2026 Main)*, \* Equal contribution, 2026.
- [6] **Haosheng Gan**, B. Tinaz, M. S. Sepehri, Z. Fabian, and M. Soltanolkotabi, “Conceptmix++: Leveling the playing field in text-to-image benchmarking via iterative prompt optimization,” in *3rd Workshop on Generative Models for Computer Vision (GMCV), CVPR*, 2025.
- [7] J. Flemings, **Haosheng Gan**, 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.

## WORK EXPERIENCE

---

### Shepherd Ventures

Financial Quant Intern & ML Team Co-lead

Apr 2024 - Oct 2024  
San Diego, CA (Hybrid)

- 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 up to **60.7%** while maintaining benchmark-competitive returns.

### Sports Media Inc

Software Engineering/AI Intern

May 2024 - Aug 2024  
Severance, CO

- Deployed production-grade fine-tuned LLMs (**Gemma**, **LLaMA**) with **distributed microservices**, enabling seamless website integration serving **6,000+** users.
- Architected a real-time **AI voice interaction** system integrating **Twilio Media Streams** with custom LLM endpoints for dynamic conversation processing.
- Designed and implemented a **real-time content recommendation system** using **collaborative filtering** and **LLM embeddings**, improving **time-on-page by 40%**.

## ACHIEVEMENTS

---

**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

**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 - Spring 2025

## ACADEMIC APPOINTMENTS

---

### Math 125: Calculus I

Course Grader

Spring 2025

University of Southern California

### Math 430: Theory of Numbers

Teaching Assistant

Fall 2024

University of Southern California

### Math 126: Calculus II

Course Grader

Spring 2024

University of Southern California

### Math 226: Calculus III

Course Grader

Fall 2023

University of Southern California