# TIANYI LORENA YAN

⊠ lorena.tianyi.yan@gmail.com 🔰 @LorenaYannnn () https://tianyi-lorena-yan-me.web.app/

I am interested in developing more factual, robust, and efficient foundation models.

#### **EDUCATION**

## University of Southern California (USC)

- **B.S.** in Computer Science
- GPA: 3.99/4.00
- Related courses: (\* Indicate graduate level course) Introduction to Artificial Intelligence (A) Introduction to Machine Learning (A) Advanced Topics in NLP (A) \*History of Language and Computing (A)

Linear Algebra and Diff. Equations (A) Probability Theory (A) Mathematical Statistics (A) Mathematics of Machine Learning (A)

## **RESEARCH EXPERIENCE & PUBLICATIONS**

#### LLMs Combine Knowledge Recall and Copy Suppression to Answer One-to-Many Factual Queries

Supervisor: Prof. Robin Jia

- Proposed Token Lens to decode attention outputs of important tokens and examined how large language models (LLMs) perform subject promotion, multi-answer retrieval, and copy suppression
- Designed and implemented causal interventions on attention outputs to demonstrate how MLPs use subject information from attention to retrieve answers and previous answer tokens for knowledge retrieval and suppression

#### **Contrastive Instruction Tuning**

Supervisor: Prof. Muhao Chen

- Proposed to leverage contrastive learning to enhance LLMs' robustness to instruction perturbation by maximizing the similarity among hidden representations of semantically equivalent instruction-input pairs
- Consistently improved LLMs' performance to perturbations in instructions across character, word, sentence, and semantic levels with +2.5% in accuracy compared to LLMs trained with instruction data augmentation

#### Monotonic Paraphrasing Improves Generalization of Language Model Prompting

Supervisor: Prof. Muhao Chen

- Implemented ensemble decoding between paraphraser and target LLMs to rephrase prompts to have low perplexity for the target models
- Evaluated and enhanced LLMs' performance on zero-shot prompting and perturbed, unseen task instructions

#### **Robust Natural Language Understanding with Residual Attention Debiasing**

Supervisor: Prof. Muhao Chen

- Implemented Product-of-Experts and residual attention learning that assembles output logits and low-layer attention scores from auxiliary model and target models to mitigate unknown biases in NLU model attention patterns
- Enhanced models' performance on out-of-distribution datasets (HANS, FEVER-Symmetric, PAWS) with improvements of 12.9%, 11.0%, and 2.7%, respectively

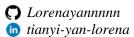
#### MuirBench: A Comprehensive Benchmark for Robust Multi-image Understanding

Supervisor: Prof. Muhao Chen

- Curated datasets to evaluate models' multi-image understanding ability across 12 tasks (scene ordering, temporal reasoning, etc.)
- Conducted human evaluation to assess model performance and validated dataset quality for benchmarking

## **PUBLICATIONS**

- Yan, T.L., & Jia, R. (2025). LLMs Combine Knowledge Recall and Copy Suppression to Answer One-to-Many Factual Queries. To be submitted to ACL 2025.
- Yan, T.L., Wang, F., Huang, J. Y., Zhou, W., Yin, F., Galstyan, A., Yin, W., & Chen, M. (2024). Contrastive Instruction Tuning. ACL 2024 Findings. PDF
- Liu, Q., Wang, F., Xu, N., Yan, T., Meng, T., & Chen, M. (2024). Monotonic Paraphrasing Improves Generalization of Language Model Prompting. EMNLP 2024 Findings. PDF



ACL 2024 Findings

#### ACL 2023 Findings

Under submission

Aug 2020 - May 2024

EMNLP 2024 Findings

To be submitted to ACL 2025

- Wang, F.\*, Huang, J. Y.\*, Yan, T., Zhou, W., & Chen, M. (2023). Robust Natural Language Understanding with Residual Attention Debiasing. ACL 2023 Findings. PDF
- Wang, F., Fu, X., ..., Yan, T., ... & Chen, M. (2024). MuirBench: A Comprehensive Benchmark for Robust Multiimage Understanding. Paper under submission. PDF

## **INDUSTRY EXPERIENCE**

Tsinghua AIR, Diffusion-Based Molecule Generation Research Intern

Supervisor: Prof. Hao Zhou

Implemented an end-to-end pipeline to jointly train EGNN-based variational autoencoder (VAE) and diffusion models for generating high-affinity ligands given protein pockets

May 2023 - Aug 2023

Jun 2022 - Aug 2022

• Pretrained unconditional VAEs on large-scale ligand-only datasets and conditioned diffusion models on protein pockets to generate ligands from the VAE's latent space, addressing the scarcity of paired ligand-pocket data

#### Microsoft M365 Deployment, Software Engineer Intern

- Designed and implemented a dashboard to centralize global deployment issues from scattered alert emails and visualize them in real time using React
- Monitored and stored real-time issue data to Cosmos NoSQL database with C# and ASP.NET

#### **AWARDS & HONORS**

•	USC Viterbi School of Engineering CURVE Research Fellowship	2022 - 2023
٠	USC Academic Achievement Award (3.75 cGPA or higher)	2022 - 2023
٠	USC ABC Innovation First Prize (1 <sup>st</sup> place among 60+ teams)	2021 - 2022
	Co-founded Ctrl+F, a platform for centralizing on-campus internship and research resources; awarded \$1,000.	

#### **TEACHING**

<ul> <li>Teaching Assistant, University of Southern California – Viterbi School of Engineering</li> <li>CSCI 467: Intro to Machine Learning</li> <li>CSCI 270: Intro to Algorithms and Theory of Computations</li> </ul>	Aug 2023 - May 2024 Jan 2022 - May 2023		
SERVICE			
Associations for Computational Linguistics / ARR (Emergency) Reviewer	2024		
Leader of USC CSSA Career Development Mentorship Program	2021 - 2023		
Organizer of 2022 California Chinese Entrepreneurship Conference	2022		

Organizer of 2022 California Chinese Entrepreneurship Conference

#### **INTERESTS & SKILLS**

- Language: Mandarin, English
- Skills: Full-stack web development, Figma
- Hobbies: Cooking, Flute, Piano, Volleyball •

Cooking is my favorite. Welcome to checkout my journal! :D