

CONTACT INFORMATION	768 Cypress Walk Goleta, CA, 93117	weizhiwang@ucsb.edu [Personal Website] [Github Repos]
RESEARCH FOCUS	<p>I am focusing on the field of Natural Language Processing. My research focus and past experience can be summarized as the following aspects:</p> <ul style="list-style-type: none"> • Multimodal Large Language Models: Augmenting LLMs with visual-knowledge (VaLM) • Long-Context Large Language Models: Long-context LLMs with memory-cache augmentation (LongMem) • Data-Centric Research: Filtering and Demistifying the pre-training and instruction-tuning data for LLMs and Multimodal LLMs (MLM-Filter) <p>Additionally, I also worked on some application areas like Multilingual Machine Translation and Speech-to-Text Translation (EMNLP'21, AAAI'22, EMNLP'22); Dialogue Systems and AI Chatbot Agents (SIGIR'22, Alexa SocialBot).</p>	
EDUCATION	<p>University of California Santa Barbara, CA, USA <i>Ph.D. Computer Science</i> Advisor: Prof. Xifeng Yan</p> <p>Rutgers University, NJ, USA <i>M.S. Computer Science</i></p> <p>Xi'an Jiaotong University, Shaanxi, China <i>B.E. Electrical Engineering</i></p>	<p>Sep 2022 – Jun 2025</p> <p>Aug 2019 – Jun 2021</p> <p>Aug 2015 – Jun 2019</p>
PREPRINTS	<p>[1] Finetuned Multimodal Language Models Are High-Quality Image-Text Data Filters Weizhi Wang, Khalil Mrini, Linjie Yang, Sateesh Kumar, Yu Tian, Xifeng Yan, Heng Wang [arXiv] [project website] [code] [model]</p> <p>[2] GPT-4V(ision) as a Generalist Evaluator for Vision-Language Tasks Xinlu Zhang*, Yujie Lu*, Weizhi Wang*, An Yan, Jun Yan, Lianke Qin, Heng Wang, Xifeng Yan, William Y. Wang, Linda R. Petzold. (*Equal Contribution) [arXiv]</p>	
PUBLICATIONS (TIME ORDER)	<p>[1] Bot or Human? Detecting ChatGPT Imposters with A Single Question Hong Wang, Xuan Luo, Weizhi Wang, Xifeng Yan. COLM 2024. [arXiv]</p> <p>[2] Augmenting Language Models with Long-Term Memory. Weizhi Wang, Li Dong, Hao Cheng, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, Furu Wei. NeurIPS 2023. [arXiv] [code(739★)]</p> <p>[3] GauchoChat: Towards Proactive, Controllable, and Personalized Social Conversation. Hong Wang, Weizhi Wang, Rajan Saini, Marina Zhukova, Xifeng Yan. Amazon Proceedings 2023. [Paper]</p> <p>[4] Visually-Augmented Language Modeling. Weizhi Wang, Li Dong, Hao Cheng, Haoyu Song, Xiaodong Liu, Xifeng Yan, Jianfeng Gao, Furu Wei. ICLR 2023. [arXiv] [code]</p>	

- [5] Non-Parametric Domain Adaptation for End-to-End Speech Translation.
Yichao Du*, **Weizhi Wang***, Zhirui Zhang, Boxing Chen, Jun Xie, Tong Xu.
EMNLP 2022. **Oral Presentation, TOP 3%**. (*Equal Contribution) [arXiv] [code]
- [6] Task-Oriented Dialogue System as Natural Language Generation.
Weizhi Wang, Zhirui Zhang, Junliang Guo, Yinpei Dai, Boxing Chen, and Weihua Luo.
SIGIR 2022. [arXiv] [code]
- [7] Regularizing End-to-End Speech Translation with Triangular Decomposition Agreement.
Yichao Du, Zhirui Zhang, **Weizhi Wang**, Boxing Chen, Jun Xie, Tong Xu, and Weihua Luo.
AAAI 2022. [arXiv] [code]
- [8] Rethinking Zero-shot Neural Machine Translation: From a Perspective of Latent Variables.
Weizhi Wang, Zhirui Zhang, Yichao Du, Boxing Chen, and Weihua Luo.
EMNLP Findings 2021. [arXiv] [code]

COMPETITIONS	<p>The Fifth Amazon Alexa Prize SocialBot Grand Challenge Nov 2022 - Sep 2023</p> <p>Champion 🏆, winning \$250,000 first place bonus</p> <p><u>Description:</u> As a member of <i>GaUCHoChat</i> team, we developed a proactive, controllable, and personalized social chatbot.</p>
PROFESSIONAL SERVICES	<p>– Conference Reviewer: KDD 2023, NeurIPS 2023, ICLR 2024, SDM 2024, ACL Rolling Review, ICML 2024, NeurIPS 2024.</p> <p>– Journal Reviewer: TASLP, TALLIP.</p>
RESEARCH INTERN EXPERIENCES	<p>Amazon, Palo Alto, US Jun 2024 – Sep 2024</p> <p><i>Mentor: Dr. Rongmei Lin</i></p> <p>Stores Foundational AI</p> <p><u>Description:</u> Vision-Language Continue Training for Large Language Models.</p> <p>TikTok, San Jose, US Sep 2023 – Dec 2023</p> <p><i>Mentor: Dr. Heng Wang</i></p> <p>Intelligent Creation Group</p> <p><u>Description:</u> Vision-language pre-training with pruned high-quality dataset.</p> <p>Microsoft Research Asia, Beijing, CN Nov 2021 – May 2023</p> <p><i>Mentor: Dr. Li Dong, Dr. Hao Cheng, Dr. Xiaodong Liu</i></p> <p>Nature Language Computing Group</p> <p><u>Description:</u> Large Language Models with visual-knowledge enhancement, Augmenting Large Language Models with long-term memory.</p> <p>Alibaba DAMO Academy, Hangzhou, CN Sep 2020 – Nov 2021</p> <p><i>Mentor: Dr. Boxing Chen, Dr. Zhirui Zhang</i></p> <p>Machine Translation Group, Language Technology Lab</p> <p><u>Description:</u> Task-Oriented Dialogue Systems, Multilingual Neural Machine Translation, and Speech-to-Text Translation.</p>
AWARDS	<p>– "Stars of Tomorrow" Excellent Intern Award, Microsoft Research, 2023.</p> <p>– Academic Excellence Fellowship. University of California Santa Barbara. 2022.</p> <p>– Outstanding Graduate Academic Performance Award. Rutgers University. 2021.</p> <p>– Siyuan Scholarship. Xi'an Jiaotong University. 2016 & 2018.</p>