Gang Li

302-367-6681 | leemagpie@gmail.com | LinkedIn | Google Scholar

Experience

Research Scientist Sep. 2024 – Present
Orby AI (acquired by Uniphore) *Mountain View, CA*

- Distillation from strong teacher models to open source models using browser-use trajectories. Obtained previous SOTA performance among open source models on WebArena benchmark.
- Learn from environment feedback with Reinforcement Learning with Verifiable Rewards (RLVR) for multi-step tasks. Improved model performance on real-world tasks with RLVR using synthetic data (paper release soon).
- Lead development of Orby's post-training infra, which supports multi-node SFT and offline/online RLVR (PPO / GRPO). Stable training with 3B-72B multimodal models and Ray-based browser environments.

Senior Software Engineer Jan. 2020 – Sep. 2024
Google Research / DeepMind *Mountain View, CA*

- Created the first foundational UI understanding model, with state-of-the-art performance on various UI tasks.
- Co-created the first unified user modeling model on visual content, with state-of-the-art performance on gaze heatmaps, aesthetics scores and human scanpaths on natrual images, UIs and designs. Extended the model to predict rich human feedback on generative images to improve text-to-image generation
- Launched models in multiple internal and user-facing products including Android Voice Access, UI-driven Pixel Testing and mobile agents, Launched models in internal products that improved user modeling and rich human feedback for Google Search and text-to-image generative models via LHF
- Mentored the team on high-quality codebase for fast iterations, downstream finetuning for diverse product requirements, and stable/scalable model serving.

Software Engineer March 2018 – Jan 2020
Google Health *Mountain View, CA*

- Created audio de-identification models and dataset for doctor-patient conversation
- Created entity and relation extraction models for doctor-patient conversation understanding

Technical Skills

Modeling: Large vision-language model pretraining and post-training, GUI agents, user modeling on visual content, entity recognition, relation extraction
Data: Crawling from Web and emulations, large-scale high-quality pretraining data generation via crawling and synthetic data, LLM-based UI interactions, large-scale user task mining from Web/mobile interaction traces
Frameworks: PyTorch, Jax, Ray, Apache Beam, Amazon MTurk
Languages: Python, C++, Javascript, HTML/CSS, Java, SQL

Education

University of Delaware Newark, DE
Ph.D. in Computer Science *Aug. 2011 – March 2018*

Beijing Institute of Technology Beijing, China
Bachelor of Engineering in Information Engineering *Aug. 2007 – May 2011*

Selected Publication

[1] **Gang Li** and Yang Li. Spotlight: Mobile UI understanding using vision-language models with a focus. In *The Eleventh International Conference on Learning Representations*, 2023. [Paper] [Google Research Blog].

[2] Youwei Liang*, Junfeng He*, **Gang Li***, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, Junjie Ke, Krishnamurthy Dj Dvijotham, Katie Collins, Yiwen Luo, Yang Li, Kai J Kohlhoff, Deepak Ramachandran, and Vidhya Navalpakkam. Rich human feedback for text-to-image generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024.
*** Co-first authors**. **Best Paper Award**. [Paper][Google Research Blog].

[3] Peizhao Li*, Junfeng He*, **Gang Li***, Rachit Bhargava, Shaolei Shen, Nachiappan Valliappan, Youwei Liang, Hongxiang Gu, Venky Ramachandran, Golnaz Farhadi, Yang Li, Kai J Kohlhoff, and Vidhya Navalpakkam. Uniar: Unifying human attention and response prediction on visual content. *NeurIPS 2024*. *** Co-first authors**. [Paper][Google Research Blog].

[4] Yang Li*, **Gang Li***, Luheng He, Jingjie Zheng, Hong Li, and Zhiwei Guan. Widget captioning: Generating natural language description for mobileuser interface elements. In *EMNLP 2020: The 2020 Conference on Empirical Methods in Natural Language Processing*, 2020. *** Co-first authors**. [Paper].

[5] **Gang Li**, Gilles Baechler, Manuel Tragut, and Yang Li. Learning to denoise raw mobile UI layouts for improving datasets at scale. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2022. [Paper].

[6] Xi Chen and 45 authors including **Gang Li**. PaLI-X: On scaling up a multilingual vision and language model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [Paper] [Google Research Blog].

[7] Bryan Wang, **Gang Li**, and Yang Li. Enabling conversational interaction with mobile UI using large language models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2023. [Paper] [Google Research Blog].

[8] Bryan Wang, **Gang Li**, Xin Zhou, Zhourong Chen, Tovi Grossman, and Yang Li. Screen2words: Automatic mobile UI summarization with multimodal learning. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 498–510, 2021. [Paper].

[9] Forrest Huang, **Gang Li**, Tao Li, and Yang Li. Automatic macro mining from interaction traces at scale. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–16, 2024. [Paper].

[10] Tao Li, **Gang Li**, Zhiwei Deng, Bryan Wang, and Yang Li. A zero-shot language agent for computer control with structured reflection. In *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023. [Paper].

[11] Mingyuan Zhong, **Gang Li**, Peggy Chi, and Yang Li. Helpviz: automatic generation of contextual visual mobile tutorials from text-based instructions. In *The 34th Annual ACM Symposium on User Interface Software and Technology*, pages 1144–1153, 2021. [Paper].

[12] Peng Su*, **Gang Li***, Cathy Wu, and K Vijay-Shanker. Using distant supervision to augment manually annotated data for relation extraction. *PloS one*, 14(7):e0216913, 2019. *** Co-first authors** [Paper].