

Bing Liu

v.bingliu@gmail.com
+1 (650) 215-8735
<http://bingliu.me>

Summary

AI research lead with deep expertise in LLM post-training, reasoning, agents, and evaluation. Proven track record of architecting and shipping GenAI models and products at Scale AI and Meta (Llama 3, Meta AI Assistant), while providing technical leadership to build high-performing research teams. Partnered with leading AI labs at Scale AI to advance foundation models and AI products from research to production. Formerly at Google Research. Author of 40+ peer-reviewed papers. Adjunct Professor at UC Santa Cruz.

Professional Experience

Scale AI Head of Research / Director - GenAI 2024 - present

- Led multi-\$M data roadmap across text, multimodal, reasoning, and agentic capabilities.
- Drove eval strategies for leading AI labs. Launched new LLM leaderboards (e.g. Humanity's Last Exam, SWE-Bench Pro) and new eval platforms adopted by top AI labs.
- Built a team of 40+ research scientists and engineers to advance LLM data and post-training research.
- Advised executive leadership on GenAI research direction and product roadmap.

Meta, Generative AI Applied Research Lead - LLM/Post-training 2023 - 2024

- Led Llama3 multilingual post-training and eval with a team of research scientists and engineers.
- Drove product data flywheel and post-training for Meta AI i18n. Grew MAU from <100M to 700M.
- Built and led a ML team launching Meta's first Llama-powered Business AI Agent at Connect 2023.

Meta, Reality Labs Tech Lead and Manager - NLP/AI 2018 - 2023

- Launched the o-to-1 self-learning system for Assistant personalization with top-line product impact.
- Led cross-functional Assistant product quality programs across ML, engineering, data, and product.
- Built the NLP team that shipped advanced AI across product verticals, languages, and devices.

Google Research Research - NLP/AI 2017

- Conducted foundational research in deep RL for spoken dialog systems.

Capio.ai (Acq. by Twilio) Software Engineer - Full Stack 2014 - 2016

- Built speech infrastructure and APIs. Developed core NLU systems for Capio's Voice AI products.

Oracle Member of Technical Staff - Big Data 2011 - 2013

- Delivered the 1st Oracle Big Data Appliance solution in APAC to a major telecom client.

Academic Affiliations

University of California, Santa Cruz Adjunct Professor 2024 - present

- Affiliated with the Computer Science and Engineering department. Gave seminars on LLMs and NLP. Guided graduate research in LLM, and multimodal AI.

Education

Carnegie Mellon University

2013 - 2018

- Ph.D., Electrical and Computer Engineering; Research area: NLP, dialog systems, deep learning
- Thesis: *Learning Task-Oriented Dialog with Neural Network Methods*

Nanyang Technological University, Singapore

2007 - 2011

- B.Eng., Electrical and Electronic Engineering; Minor Degree in Business
- Exchange program at KTH Royal Institute of Technology, Sweden

Professional Service

- Publication Chair: ACL 2024
- Area Chair: Large Language Models track @ ACL 2023, Senior PC: ACL, EMNLP, AACL, ICASSP
- General Chair: NLP for Conversational AI Workshop @ ACL 2022
- Program Committee: ICLR, ACL, EMNLP, NAACL, ICASSP, Interspeech, AACL, etc
- Journal Reviewer: IEEE TASLP, IEEE SPL, Neural Networks, Info Systems

Selected Publications

Published 40+ papers in ICLR, ACL, EMNLP, NAACL, AACL, etc. Full list: [Google Scholar](#).

1. Mazeika, Mantas, et al. "**Remote Labor Index: Measuring AI Automation of Remote Work**", in arXiv, 2025
2. Xiang Deng, Jeff Da, Edwin Pan, Yannis Y. He, Charles Ide, Kanak Garg, Niklas Lauffer, Andrew Park, Chetan Rane, Karmini Sampath, Maya Krishnan, Srivatsa R Kundurthy, Sean M. Hendryx, Zifan Wang, Chen Bo Calvin Zhang, Noah Jacobson, Bing Liu, Brad Kenstler, "**SWE-Bench Pro: Can AI Agents Solve Long-Horizon Software Engineering Tasks?**", in arXiv, 2025
3. Xingang Guo, Utkarsh Tyagi, Advait Gosai, Paula Vergara, Ernesto Gabriel Hernández Montoya, Chen Bo Calvin Zhang, Bin Hu, Yunzhong He, Bing Liu, Rakshith Sharma Srinivasa, "**Beyond Seeing: Evaluating Multimodal LLMs on Tool-Enabled Image Perception, Transformation, and Reasoning**", in arXiv, 2025
4. Junkai Zhang, Zihao Wang, Lin Gui, Swarnashree Mysore Sathyendra, Jaehwan Jeong, Victor Veitch, Wei Wang, Yunzhong He, Bing Liu, Lifeng Jin, "**Chasing the Tail: Effective Rubric-based Reward Modeling for Large Language Model Post-Training**", in arXiv, 2025
5. Anisha Gunjal, Anthony Wang, Elaine Lau, Vaskar Nath, Yunzhong He, Bing Liu, and Sean Hendryx, "**Rubrics as Rewards: Reinforcement Learning Beyond Verifiable Domains**", in arXiv, 2025
6. Alexander R. Fabbri, Diego Mares, Jorge Flores, Meher Mankikar, Ernesto Hernandez, Dean Lee, Bing Liu, Chen Xing, "**MultiNRC: A Challenging and Native Multilingual Reasoning Evaluation Benchmark for LLMs**", in arXiv, 2025
7. Lucas Bandarkar, Benjamin Muller, Pritish Yuvraj, Rui Hou, Nayan Singhal, Hongjiang Lv, Bing Liu, "**Layer Swapping for Zero-Shot Cross-Lingual Transfer in Large Language Models**", in ICLR, 2024
8. Dubey, Abhimanyu, et al, "**The Llama 3 Herd of Models**", in arXiv, 2024

9. Derek Xu, Shuyan Dong, Changhan Wang, Suyoun Kim, Zhaojiang Lin, Akshat Shrivastava, Shang-Wen Li, Liang-Hsuan Tseng, Alexei Baevski, Guan-Ting Lin, Bing Liu, Hung-yi Lee, Yizhou Sun and Wei Wang, "**Introducing Semantics into Speech Encoders**", in ACL, 2023
10. Zhiyu Chen, Bing Liu, Seungwhan Moon, Chinnadhurai Sankar, Paul Crook, William Yang Wang, "**KETOD: Knowledge-Enriched Task-Oriented Dialogue**", in NAACL, 2022
11. Zhaojiang Lin, Bing Liu, Andrea Madotto, Seungwhan Moon, Zhenpeng Zhou, Paul Crook, Zhiguang Wang, Zhou Yu, Eunjoon Cho, Rajen Subba, Pascale Fung, "**Zero-Shot Dialogue State Tracking via Cross-Task Transfer**", in EMNLP, 2021
12. Andrea Madotto, Zhaojiang Lin, Zhenpeng Zhou, Seungwhan Moon, Paul Crook, Bing Liu, Zhou Yu, Eunjoon Cho, Zhiguang Wang, "**Continual Learning in Task-Oriented Dialogue Systems**", in EMNLP, 2021
13. Zhiyu Chen, Honglei Liu, Hu Xu, Seungwhan Moon, Hao Zhou, Bing Liu, "**NUANCED: Natural Utterance Annotation for Nuanced Conversation with Estimated Distributions**", in EMNLP, 2021
14. Zhaojiang Lin, Bing Liu, Seungwhan Moon, Paul Crook, Zhenpeng Zhou, Zhiguang Wang, Zhou Yu, Andrea Madotto, Eunjoon Cho, Rajen Subba, "**Leveraging Slot Descriptions for Zero-Shot Cross-Domain Dialogue State Tracking**", in NAACL, 2021
15. Kai Sun, Seungwhan Moon, Paul Crook, Stephen Roller, Becka Silvert, Bing Liu, Zhiguang Wang, Honglei Liu, Eunjoon Cho, and Claire Cardie, "**Adding Chit-Chats to Enhance Task-Oriented Dialogues**", in NAACL, 2021
16. Tianxing He, Jun Liu, Kyunghyun Cho, Myle Ott, Bing Liu, James Glass, Fuchun Peng, "**Analyzing the Forgetting Problem in the Pretrain-Finetuning of Dialogue Response Models**", in EACL, 2021
17. Hu Xu, Seungwhan Moon, Honglei Liu, Bing Liu, Pararth Shah, Philip S Yu, "**User memory reasoning for conversational recommendation**", in COLING, 2020

Patents

1. "**Personalized conversational recommendations by assistant systems**", US Patent, April 2023
2. "**Dialog Session Override Policies for Assistant Systems**", US Patent, May 2022
3. "**Systems and Methods for Implementing Smart Assistant Systems**", US Patent, April 2022

Skills

- LLMs, agentic systems, reasoning, RLHF, evaluation, multimodal AI, Python, PyTorch, product strategy, team building, roadmap planning, research leadership