

Hexiang (Frank) Hu

hexiang.frank.hu@gmail.com
https://www.hexianghu.com

EMPLOYMENT	Member of Technical Staff	Oct 2024 - Present
	SpaceXAI (formerly xAI)	Redmond, WA
	Research Scientist	Mar 2023 - Oct 2024
	Google DeepMind	Seattle, WA
	Research Scientist	Jun 2021 - Mar 2023
	Google Brain	Seattle, WA
EDUCATION	University of Southern California	2017 - 2021
	<i>Ph.D.</i> in Computer Science	
	University of California, Los Angeles (UCLA)	2016 - 2017
	<i>Ph.D.</i> in Computer Science (Transfer)	
	Simon Fraser University	2013 - 2016
	<i>Bachelor of Science</i> in Computer Science	
	Zhejiang University	2011 - 2016
	<i>Bachelor of Engineering</i> in Computer Science and Technology	
PROJECTS	Grok Multimodal	Jun 2026 - Present
	SpaceXAI <i>Multimodal Lead</i>	
	• Lead Grok 4.6 multi-modal data and training	Aug 2026
	— #3 on the AA Intelligence Index at launch	
	• Lead Grok 4.5 multi-modal data and training	Jul 2026
	— #4 on the AA Intelligence Index at launch, with best token efficiency	
	Grok Imagine	Aug 2025 - Jul 2026
	SpaceXAI <i>Image Lead</i>	
	• Lead Imagine Image 2.0 pre-training	Aug 2026
	— #2 on Image & Image Edit Arenas at launch	
	• Lead Grok Imagine v1.5 (quality mode)	May 2026
	— top-3 on Image Edit Arena at launch	
	• Lead Grok Imagine v0.9 & v1.0 (image & video)	Sep 2025 - Feb 2026
	— #1 on Video & #6 on Image Arena at launch	
	Grok Native Image Generation	Dec 2024 - Aug 2025
	SpaceXAI <i>Lead</i>	
	• Lead Grok native text-to-image and image-to-image model	Feb 2025
	— grew DAU from $O(100K)$ to $O(10M)$	
	• Core Contributor to Grok text-to-image model (Aurora)	Dec 2024
	Gemini & Imagen	2023 - 2024
	Google DeepMind <i>Core Contributor</i>	
	• Gemini 2.0 native image generation	Dec 2024
	• Imagen 3 post-training (Instruct-Imagen)	May 2024
	• Gemini multi-modal fine-tuning (MMFT)	2023 - 2024

*: Authors contributed equally

PUBLICATIONS

Lichang Chen, **Hexiang Hu**, Mingda Zhang, Yiwen Chen, Zifeng Wang, Yandong Li, Pranav Shyam, Tianyi Zhou, Heng Huang, Ming-Hsuan Yang, and Boqing Gong. “OmnixR: Evaluating Omni-modality Language Models on Reasoning across Modalities”. In: *ICLR*. 2025

Jiacheng Chen, Tianhao Liang, Sherman Siu, Zhengqing Wang, Kai Wang, Yubo Wang, Yuansheng Ni, Wang Zhu, Ziyang Jiang, Bohan Lyu, Dongfu Jiang, Xuan He, Yuan Liu, **Hexiang Hu**, Xiang Yue, and Wenhui Chen. “MEGA-Bench: Scaling Multimodal Evaluation to over 500 Real-World Tasks”. In: *ICLR*. 2025

Nanye Ma, Shangyuan Tong, Haolin Jia, **Hexiang Hu**, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, and Saining Xie. “Inference-Time Scaling for Diffusion Models beyond Scaling Denoising Steps”. In: *CVPR (Highlight)*. 2025

Kai Zhang, Yi Luan, **Hexiang Hu**, Kenton Lee, Siyuan Qiao, Wenhui Chen, Yu Su, and Ming-Wei Chang. “MagicLens: Self-Supervised Image Retrieval with Open-Ended Instructions”. In: *ICML (Oral)*. 2024

Hexiang Hu^{*}, Kelvin CK Chan^{*}, Yu-Chuan Su^{*}, Wenhui Chen^{*}, Yandong Li, Kihyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William Cohen, et al. “Instruct-Imagen: Image generation with multi-modal instruction”. In: *CVPR (Oral)*. 2024

Cong Wei, Yang Chen, Haonan Chen, **Hexiang Hu**, Ge Zhang, Jie Fu, Alan Ritter, and Wenhui Chen. “UniIR: Training and benchmarking universal multimodal information retrievers”. In: *ECCV (Oral)*. 2024

Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, et al. “Pali-x: On scaling up a multilingual vision and language model”. In: *arXiv preprint arXiv:2305.18565* (2023)

Peter Shaw, Mandar Joshi, James Cohan, Jonathan Berant, Panupong Pasupat, **Hexiang Hu**, Urvashi Khandelwal, Kenton Lee, and Kristina N Toutanova. “From Pixels to UI Actions: Learning to Follow Instructions via Graphical User Interfaces”. In: *NeurIPS (Spotlight)*. 2023

Wenhui Chen^{*}, **Hexiang Hu**^{*}, Yandong Li, Nataniel Ruiz, Xuhui Jia, Ming-Wei Chang, and William W Cohen. “Subject-driven Text-to-Image Generation via Apprenticeship Learning”. In: *NeurIPS*. 2023

Kenton Lee, Mandar Joshi, Iulia Turc, **Hexiang Hu**, Fangyu Liu, Julian Eisenschlos, Urvashi Khandelwal, Peter Shaw, Ming-Wei Chang, and Kristina Toutanova. “Pix2Struct: Screenshot Parsing as Pretraining for Visual Language Understanding”. In: *ICML (Oral)*. 2023

Yang Chen, **Hexiang Hu**, Yi Luan, Haitian Sun, Soravit Changpinyo, Alan Ritter, and Ming-Wei Chang. “Can Pre-trained Vision and Language Models Answer Visual Information-Seeking Questions?” In: *EMNLP*. 2023

Hexiang Hu, Yi Luan, Yang Chen, Urvashi Khandelwal, Mandar Joshi, Kenton Lee, Kristina Toutanova, and Ming-Wei Chang. “Open-domain Visual Entity Recognition: Towards Recognizing Millions of Wikipedia Entities”. In: *ICCV (Oral)*. 2023

Jihyung Kil, Soravit Changpinyo, Xi Chen, **Hexiang Hu**, Sebastian Goodman, Wei-Lun Chao, and Radu Soricut. “Prestu: Pre-training for scene-text understanding”. In: *ICCV*. 2023

Wenhu Chen, **Hexiang Hu**, Chitwan Saharia, and William W Cohen. “Re-imagined: Retrieval-augmented text-to-image generator”. In: *ICLR*. 2023

Hexiang Hu, Ozan Sener, Fei Sha, and Vladlen Koltun. “Drinking from a Firehose: Continual Learning with Web-scale Natural Language”. In: *T-PAMI* (2023)

Wenhu Chen, **Hexiang Hu**, Xi Chen, Pat Verga, and William W Cohen. “Murag: Multimodal retrieval-augmented generator for open question answering over images and text”. In: *EMNLP*. 2022

Tai-Yu Pan, Cheng Zhang, Yandong Li, **Hexiang Hu**, Dong Xuan, Soravit Changpinyo, Boqing Gong, and Wei-Lun Chao. “On model calibration for long-tailed object detection and instance segmentation”. In: *NeurIPS*. 2021

Linlu Qiu, **Hexiang Hu**, Bowen Zhang, Peter Shaw, and Fei Sha. “Systematic Generalization on gSCAN: What is Nearly Solved and What is Next?”. In: *EMNLP*. 2021

Bowen Zhang, **Hexiang Hu**, Linlu Qiu, Peter Shaw, and Fei Sha. “Visually Grounded Concept Composition”. In: *Findings of EMNLP*. 2021

Cheng Zhang, Tai-Yu Pan, Yandong Li, **Hexiang Hu**, Dong Xuan, Soravit Changpinyo, Boqing Gong, and Wei-Lun Chao. “MosaicOS: a simple and effective use of object-centric images for long-tailed object detection”. In: *ICCV*. 2021

Han-Jia Ye, **Hexiang Hu**, and De-Chuan Zhan. “Learning adaptive classifiers synthesis for generalized few-shot learning”. In: *IJCV*. Springer, 2021

Jiacheng Chen*, **Hexiang Hu***, Hao Wu, Yuning Jiang, and Changhu Wang. “Learning the Best Pooling Strategy for Visual Semantic Embedding”. In: *CVPR (Oral)*. 2021

Bowen Zhang*, **Hexiang Hu***, Vihan Jain, Eugene Ie, and Fei Sha. “Learning to Represent Image and Text with Denotation Graph”. In: *EMNLP*. 2020

Wang Zhu*, **Hexiang Hu***, Jiacheng Chen, Zhiwei Deng, Vihan Jain, Eugene Ie, and Fei Sha. “BabyWalk: Going Farther in Vision-and-Language Navigation by Taking Baby Steps”. In: *ACL*. 2020

Han-Jia Ye, **Hexiang Hu**, De-Chuan Zhan, and Fei Sha. “Few-shot learning via embedding adaptation with set-to-set functions”. In: *CVPR*. 2020

Risto Vuorio*, Shao-Hua Sun*, **Hexiang Hu**, and Joseph J Lim. “Multimodal Model-Agnostic Meta-Learning via Task-Aware Modulation”. In: *NeurIPS (Spotlight)*. 2019

Hexiang Hu, Ishan Misra, and Laurens van der Maaten. “Evaluating text-to-image matching using binary image selection (BISON)”. In: *ICCV Workshop*. 2019

Bowen Zhang, **Hexiang Hu**, and Fei Sha. “Visual Storytelling via Predicting Anchor Word Embeddings in the Stories”. In: *ICCV Workshop*. 2019

Nelson Nauata, **Hexiang Hu**, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, and Greg Mori. “Structured label inference for visual understanding”. In: *T-PAMI* (2019)

Kurt Shuster, Samuel Humeau, **Hexiang Hu**, Antoine Bordes, and Jason Weston. “Engaging image captioning via personality”. In: *CVPR*. 2019

Hexiang Hu*, Liyu Chen*, Boqing Gong, and Fei Sha. “Synthesized policies for transfer and adaptation across tasks and environments”. In: *NeurIPS (Spotlight)*. 2018

Soravit Changpinyo, **Hexiang Hu**, and Fei Sha. “Multi-task learning for sequence tagging: An empirical study”. In: *COLING*. 2018

Bowen Zhang*, **Hexiang Hu***, and Fei Sha. “Cross-modal and hierarchical modeling of video and text”. In: *ECCV*. 2018

Hexiang Hu*, Wei-Lun Chao, and Fei Sha. “Learning answer embeddings for visual question answering”. In: *CVPR*. 2018

Wei-Lun Chao*, **Hexiang Hu***, and Fei Sha. “Cross-dataset adaptation for visual question answering”. In: *CVPR*. 2018

Wei-Lun Chao*, **Hexiang Hu***, and Fei Sha. “Being negative but constructively: Lessons learnt from creating better visual question answering datasets”. In: *NAACL-HLT (Oral)*. 2018

Chao-Yuan Wu, Manzil Zaheer, **Hexiang Hu**, R Manmatha, Alexander J Smola, and Philipp Krähenbühl. “Compressed video action recognition”. In: *CVPR*. 2018

Hexiang Hu*, **Shiyi Lan***, Yuning Jiang, Zhimin Cao, and Fei Sha. “Fastmask: Segment multi-scale object candidates in one shot”. In: *CVPR (Spotlight)*. 2017

Hexiang Hu, Guang-Tong Zhou, Zhiwei Deng, Zicheng Liao, and Greg Mori. “Learning structured inference neural networks with label relations”. In: *CVPR*. 2016

Zhiwei Deng, Arash Vahdat, **Hexiang Hu**, and Greg Mori. “Structure inference machines: Recurrent neural networks for analyzing relations in group activity recognition”. In: *CVPR*. 2016

TECHNICAL REPORTS

Gemini Team. “Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities”. In: *Technical Report*. 2025

Imagen 3 Team. “Imagen 3: Google’s highest quality text-to-image model”. In: *Technical Report*. 2024

Gemini Team. “Gemini: a family of highly capable multimodal models”. In: *Technical Report*. 2023

PREPRINTS

Hsin-Ping Huang, Xinyi Wang, Yonatan Bitton, Hagai Taitelbaum, Gaurav Singh Tomar, Ming-Wei Chang, Xuhui Jia, Kelvin CK Chan, **Hexiang Hu**, Yu-Chuan Su, and Ming-Hsuan Yang. “KITTEEN: A Knowledge-Intensive Evaluation of Image Generation on Visual Entities”. In: *arXiv preprint arXiv:2410.11824* (2024)

PATENTS

R Manmatha, **Hexiang Hu**, and Deva Ramanan. *Compressed content object and action detection*. US Patent 10,528,819. Jan. 2020

Vihan Jain, JoonSeok Lee, Ming Zhao, Sheide Chammas, **Hexiang Hu**, Bowen Zhang, Fei Sha, and Tze Way Eugene Ie. *Hierarchical video encoders*. US Patent 11,533,495. Dec. 2022

SERVICE

Senior Area Chair

- The North American Chapter of the ACL (NAACL) 2025

Journal Reviewer

- IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)

- International Journal of Computer Vision (IJCV)

Conference Reviewer

- Conference on Neural Information Processing Systems (NeurIPS)
- International Conference on Machine Learning (ICML)
- International Conference on Learning Representations (ICLR)
- Annual Meeting of the Association for Computational Linguistics (ACL)
- The North American Chapter of the ACL (NAACL)
- IEEE Conference on Computer Vision and Pattern Recognition (CVPR)
- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)
- Special Interest Group on Computer Graphics (SIGGRAPH)
- Asian Conference on Computer Vision (ACCV)
- Winter Conference on Applications of Computer Vision (WACV)
- AAAI Conference on Artificial Intelligence (AAAI)
- International Conference on Image Processing (ICIP)

TECHNICAL TALKS

- “Towards Unified Visual Generation and Editing”
- CVPR, Seattle, WA Jun 2024
 - Google DeepMind 2024
 - xAI 2024
 - Microsoft AI 2024
 - OpenAI 2024
- “Understanding Language in Perception and Embodiment”
- Google, Mountain View, CA Feb 2021
 - Facebook, New York, NY Feb 2021
 - Waymo, Mountain View, CA Mar 2021
 - University of Washington, Seattle, WA Feb 2021
 - University of California, Berkeley, CA Mar 2021
 - University of Michigan, Ann Arbor, MI Mar 2021
 - University of California, Santa Cruz, CA Apr 2021
- “Learning Visually Grounded Semantics with Denotation Graph”
- Google, Mountain View, CA Sep 2020
 - USC, Los Angeles, CA Mar 2021
- “Learning Generalizable Neural Agents across Different Task Horizons”
- Google, Mountain View, CA Nov 2019
- “Synthesized Policies for Transfer and Adaptation across Tasks and Environments”
- NeurIPS, Montreal, QC Dec 2018
- “Being Negative but Constructively: Lessons from Creating Better VQA Datasets”
- NAACL, New Orleans, LA Jun 2018
 - USC, Los Angeles, CA Apr 2021

TEACHING

- Teaching Assistant*, University of Southern California
- CSCI-599: Deep Learning and its Applications Fall 2017 - 2018
- CSCI-567: Machine Learning Spring 2020 - 2021

AWARDS & HONORS

Awards and Prizes

- NeurIPS Student Travel Award 2018, 2019
- Simon Fraser University President's Honour Roll 2014, 2015
- Zhejiang University First-class Academic Excellence Award 2011, 2012

Scholarship & Fellowship

- USC CS Department Fellowship 2017
- UCLA CS Department Fellowship 2016
- Simon Fraser University Entrance Scholarship 2013

PRESS COVERAGE

- “BabyWalk AI breaks complex navigation into simple steps”
by Kyle Wiggers, VentureBeat May 12, 2020
- “Facebook aims to be more engaging to humans in latest AI research”
by Tiernan Ray, ZDNet Oct 26, 2018

ADVISING & MENTORSHIP

Junior Ph.D. Mentor

- Bowen Zhang, Ph.D. student at USC (Now: AI Research Scientist at Meta)
- Liyu Chen, Ph.D. student at USC (Now: Member of Technical Staff at OpenAI)
- Shao-Hua Sun, Ph.D. student at USC (Now: Assistant Professor at National Taiwan University)

Visiting Scholar & Intern Mentor

- Willis Ma, Student Researcher at Google DeepMind (Now: Ph.D. student at NYU Courant)
- Shangyuan Tong, Research Intern at Google DeepMind (Now: Tesla AI)
- Ziyu Wan, Student Researcher at Google DeepMind
- Lichang Chen, Student Researcher at Google DeepMind (Now: Research Scientist at Meta Superintelligence Lab)
- Yingcheng Liu, Student Researcher at Google DeepMind (Now: Ph.D. candidate at MIT CSAIL)
- Kai Zhang, Student Researcher at Google DeepMind (Now: Founding Member of Technical Staff at NeoCognition)
- Cong Wei, Student Collaborator at University of Waterloo (Now: Member of Technical Staff at SpaceXAI)
- Yang Chen, Student Researcher at Google Brain (Now: Research Scientist at NVIDIA)
- Jihyung Kil, Student Researcher at Google (Now: Research Scientist at Adobe Research)
- Cheng Zhang, Student Collaborator at Ohio State University (Now: Assistant Professor at Texas A&M University)
- Wang Zhu, Visiting Student from SFU (Now: Postdoctoral Researcher at UC Berkeley)
- Jiacheng Chen, Undergraduate at SFU & Zhejiang University (Now: Alibaba Token Hub)
- Han-Jia Ye, Visiting Scholar from Nanjing University (Now: Associate Professor at Nanjing University)
- Risto Vuorio, Visiting Researcher from SK T-Brain (Now: Reflection AI, after Ph.D. at Oxford)

- Shiyi Lan, Undergraduate at Fudan University (Now: Senior Research Scientist at NVIDIA)
- Linlu Qiu, AI Resident at Google (Now: Ph.D. student at MIT CSAIL)

Undergraduate Student Mentor

- Yuna Lee, USC (Now: Software Engineer at Google)