Hexiang (Frank) Hu

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

EXPERIENCE

SpaceXAI (formerly xAI) | Member of Technical Staff, Redmond, WA — Nov 2024 - Present

Grok Multimodal Understanding — May 2026 - Present

Multimodal Data Lead (team manager)

- Curate and ablate multimodal (mid-train + SFT) data for mm agent, computer use agent, visual coding data: off-policy agent trace synthesis, thinking-trace behavior correction, etc
- Built an agentic data system that autonomously handles data planning, annotation, curation, thinking-trace generation, and quality review (w/ AI engineers in the loop for critical decision making)
- Led multi-modal mid-train and SFT for Grok 4.5 / 4.6 / 4.7 / 4.8

Grok Imagine Image — Jun 2025 - Jun 2026

Pre-train and Data Lead (team manager)

- Drove large-scale training data processing at $O(10\text{T})$ tokens, spanning curation, captioning, rebalancing, and synthesizing image pair data and interleaved image data from web-mined images, video shots, and clustered image sets
- Pre-trained a series of diffusion Transformers, from 10B dense to 200B MoE as base model for image and video generation
- Led pre-training for Grok Imagine Image v1.5 (quality mode) and v2.0
- Led Grok Imagine v0.9 and v1.0 (image pod)

Grok Native Image Generation — Dec 2024 - Jun 2025

Contributing as a team of two

- Worked end-to-end across data, pre-training, post-training, and inference
- Led Grok's native text-to-image and image-to-image model

Google DeepMind (formerly Google Brain) | Research Scientist, Seattle, WA — Jun 2021 - Nov 2024

- Core contributor to Gemini 2.0 native image generation
- Core contributor to Imagen 3 post-training (Instruct-Imagen)
- Core contributor to Gemini multi-modal fine-tuning (MMFT)

EDUCATION

University of Southern California | Ph.D. in Computer Science, advised by Prof. Fei Sha — 2017 - 2021
Simon Fraser University | B.Sc. in Computer Science — 2013 - 2016
Zhejiang University | B.Eng. in Computer Science and Technology — 2011 - 2016

SELECTED PUBLICATIONS (full list: https://www.hexianghu.com/cv/latest.pdf)

- *Gemini 2.5* (2025), *Imagen 3* (2024), and *Gemini* (2023) technical reports – core contributor
- *Inference-Time Scaling for Diffusion Models beyond Scaling Denoising Steps.* CVPR 2025 (Highlight)
- *Instruct-Imagen: Image Generation with Multi-modal Instruction.* CVPR 2024 (Oral)
- *MagicLens: Self-Supervised Image Retrieval with Open-Ended Instructions.* ICML 2024 (Oral)
- *UniIR: Training and Benchmarking Universal Multimodal Information Retrievers.* ECCV 2024 (Oral)
- *Open-domain Visual Entity Recognition.* ICCV 2023 (Oral)
- *Pix2Struct: Screenshot Parsing as Pretraining for Visual Language Understanding.* ICML 2023 (Oral)
- *From Pixels to UI Actions: Learning to Follow Instructions via Graphical User Interfaces.* NeurIPS 2023 (Spotlight)
- *Subject-driven Text-to-Image Generation via Apprenticeship Learning.* NeurIPS 2023
- *Re-Imagen: Retrieval-Augmented Text-to-Image Generator.* ICLR 2023
- *Drinking from a Firehose: Continual Learning with Web-scale Natural Language.* T-PAMI 2023
- *Learning the Best Pooling Strategy for Visual Semantic Embedding.* CVPR 2021 (Oral)
- *BabyWalk: Going Farther in Vision-and-Language Navigation by Taking Baby Steps.* ACL 2020
- *Few-Shot Learning via Embedding Adaptation with Set-to-Set Functions (FEAT).* CVPR 2020
- *Synthesized Policies for Transfer and Adaptation across Tasks and Environments.* NeurIPS 2018 (Spotlight)
- *Being Negative but Constructively: Lessons Learnt from Creating Better VQA Datasets.* NAACL 2018 (Oral)
- *FastMask: Segment Multi-scale Object Candidates in One Shot.* CVPR 2017 (Spotlight)

SERVICE & MENTORSHIP

- Senior Area Chair, NAACL 2025; reviewer for NeurIPS, ICML, ICLR, CVPR, ICCV, ECCV, ACL, T-PAMI, and IJCV
- Mentored 20 students and visiting researchers, now at OpenAI, Meta, NVIDIA, SpaceXAI, and in faculty roles