Songyao Jiang

Boston MA, USA (open to relocate) • Tel: +1(734) 546-0695
jiangsongyao@gmail.com • GitHub • Homepage • Google Scholar • LinkedIn

SUMMARY

Senior Applied Scientist at Amazon AGI, specializing in multimodal foundation models. I lead multimodal joint reinforcement learning efforts that improve cross-modal understanding and reasoning across speech, images, video, and text—helping bring Amazon Nova 1.0 and 2.0 from research to launch.

EXPERTISE

Multimodal Foundation Models: Pretrain and post-train multimodal LLMs jointly across text, image, video, and speech modalities. Design and develop reward functions/models for agentic RL tasks. Develop data pipelines to prepare high-quality data mix for cold-start SFT and agentic RL. Adapt large-scale production training on custom chips such as Trainium.

Research Interests: Multimodal generative models, reinforcement learning, computer vision.

EDUCATION

Northeastern University — **Boston, MA**
Ph.D. in Computer Engineering — **06/2016 – 05/2022**

University of Michigan — **Ann Arbor, MI**
Master of Science in Electrical and Computer Engineering — **09/2013 – 05/2015**

Hong Kong Polytechnic University — **Hong Kong**
Bachelor of Engineering in Electrical and Computer Engineering — **09/2009 – 06/2013**

WORK EXPERIENCE

Amazon.com, Inc. — **Boston, MA**

Senior Applied Scientist - AGI — **4/2025 – present**
- Led multimodal LLM post-training (RL/preference optimization) to improve reasoning and agentic capability across text, image, video, speech modalities,
- Built data pipeline to prepare multi-modal long-reasoning and tool-calling dataset.
- Improved key evaluation benchmarks (MMMU/MathVision/VQA/ASR/WER etc.)
- Resulted in the launch of Nova-family MLLMs including Nova 1.0 and 2.0.

Applied Scientist - AGI — **11/2023 – 3/2025**
- Designed multi-modal data mixture + filtering pipeline
- Enabled efficient training on AWS Trainium chips by implementing sharding strategy, resolving numerical differences, and increasing overall training throughput.

Applied Scientist - Device AI — **6/2022 – 11/2023**
- Delivered end-to-end on-device AI solutions (CV modeling, data/evaluation, optimization) for Amazon Devices such as Ring camera, Alexa Echo Show, and Kindle Fire tablets.

AInnovation Labs, Inc. — **Boston, MA**

Founding Member and Research Engineer — **1/2021 – 5/2022**
- Led core algorithm development including object detection, face recognition, and pose estimation.

Giaran, Inc. (startup) — **Boston, MA**

Founding Member and Research Engineer — **6/2016 – 11/2017**
- Developed key machine learning algorithms in AI beauty products.
- Our startup was then acquired by Shiseido Americas.

CHALLENGES

CVPR 2021 Challenge on Agriculture-Vision Pattern Recognition **04/2021 – 06/2021**

- Team leader and first contributor. Ranked the 4th place in supervised track. [GitHub][Leaderboard]
- Developed a multi-modal and self-constructing GCN for multi-label agricultural pattern recognition given RGB and infra-red aerial agriculture images.

CVPR 2021 Challenge on Signer-Independent Isolated Sign Language Recognition **12/2020 – 04/2021**

- Team leader and first contributor. **1st place winner** in both RGB and RGB+D tracks. [GitHub][Leaderboard]
- Proposed a novel spatio-temporal GCN with attention mechanism to learn dynamics in whole-body skeleton graph as well as fusing with RGB, optical flow and depth HHA video modalities via a unified skeleton-aware multi-modal framework to recognize sign language glosses from input RGB+D videos.

PUBLICATIONS

- Z. Tang, X. Zhang, **S. Jiang**, and D. Modolo, "MM-ReCoder: Advancing Chart-to-Code Generation with Reinforcement Learning and Self-Correction," *Under Review by CVPR*, 2025.
- Q. Li, J. Zhou, **S. Jiang**, Y. Fu, and M. Su, "Single-Cell Array Enhanced Cell Damage Recognition Using Artificial Intelligence for Anticancer Drug Discovery," *Analytical Chemistry 97 (7),* pp. *4202-4208*, 2025.
- B. Sun, Y. Zhang, **S. Jiang**, and Y. Fu, "Hybrid Pixel-Unshuffled Network for Lightweight Image Super-Resolution," *AAAI*, 2023. [Preprint][GitHub][Demo]
- **S. Jiang**, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, "Sign Language Recognition via Skeleton-aware Multi-modal Ensemble," *Under Review*, 2022. [Preprint][GitHub]
- **S. Jiang**, B. Sun, L. Wang, Y. Bai, K. Li, and Y. Fu, "Skeleton Aware Multi-modal Sign Language Recognition," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2021. [Paper][GitHub]
- **S. Jiang**, Z. Tao, and Y. Fu, "Geometrically Editable Face Image Translation with Adversarial Networks," *IEEE Transactions on Image Processing (TIP)*, vol. 30, pp. *2771-2783*, 2021. [Paper]
- **S. Jiang**, H. Liu, Y. Wu, and Y. Fu, "Spatially Constrained GAN for Face and Fashion Synthesis," in *16th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2021. [Paper][GitHub][Award][Web]
- Y. Yin, J. P. Robinson, **S. Jiang**, and Y. Fu, "SuperFront: From Low-resolution to High-resolution Frontal Face Synthesis," in *Proceedings of ACM Multimedia (ACMMM)*, 2021. [Paper][GitHub]
- Y. Yin, **S. Jiang**, J. P. Robinson, and Y. Fu, "Dual-attention GAN for Large-pose Face Frontalization," in *15th IEEE International Conference on Automatic Face & Gesture Recognition (FG),* 2020. [Paper][GitHub]
- S. Sarkar, W. Kang, **S. Jiang**, K. Li, S. Ray, E. Luther, A. R. Ivanov, Y. Fu, and T. Konry, "Machine Learning-aided Quantification of Antibody-based Cancer Immunotherapy by Natural Killer Cells in Microfluidic Droplets," Lab on a Chip, 20(13), pp. *2317-2327*, 2020. [Paper]
- **S. Jiang**, Z. Tao, and Y. Fu, "Segmentation Guided Image-to-Image Translation with Adversarial Networks," in *14th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2019. [Paper][GitHub]
- T. Alashkar, **S. Jiang**, and Y. Fu, "Rule-Based Facial Makeup Recommendation System," in *12th IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, 2017. [Paper]
- T. Alashkar, **S. Jiang**, S. Wang, and Y. Fu, "Examples-Rules Guided Deep Neural Network for Makeup Recommendation," in *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*, 2017. [Paper]
- **S. Jiang** and T. Kato, "Dynamic Modelling of Combined Cycle Power Plant for Load Frequency Control with Large Penetration of Renewable Energy," in *7th JUACEP Workshop*. 2014.

PATENTS

- S. Gao, J. Thomas, J. Zhang, **S. Jiang**, J. Luo, "Stylus Input Compensation System," Granted, *US Patent 12,293,040*. [Patent]
- Y. Fu, **S. Jiang**, B. Sun, "Light-Weight Pose Estimation Network with Multi-Scale Heatmap Fusion," Granted. *US Patent 12,205,317.* [Patent]
- Y. Fu, **S. Jiang**, "Segmentation Guided Image Generation with Adversarial Networks,"

Granted. *US Patent 10,825,219.* [Patent]

- Y. Fu, **S. Jiang**, "Video 2D Multi-person Pose Estimation using Multi-frame Refinement and Optimization," Published. *WIPO Patent App. No.: WO 2020/232069*. [Patent]
- Y. Fu, S. Wang, S. Lee, **S. Jiang**, B. Sun, H. Mao, K. H. E. Cheung, "Systems and Methods for Virtual Facial Makeup Removal and Simulation, Fast Facial Detection and Landmark Tracking, Reduction in Input Video Lag and … ," Published. *US Patent App. No: 16/584,310.* [Patent]

ACADEMIC SERVICE

Conference PC Member and Reviewer

- International Conference on Computer Vision (ICCV)
- European Conference on Computer Vision (ECCV)
- International Joint Conferences on Artificial Intelligence (IJCAI)
- IEEE International Conference on Automatic Face & Gesture Recognition (FG)
- IEEE International Conference on Data Mining (ICDM)
- IEEE International Conference on Multimedia Information Processing and Retrieval (MIPR)

Journal Reviewer

- IEEE Transactions on Image Processing (TIP)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Transactions on Multimedia (TMM)
- Journal of Visual Communication and Image Representation (JVCI)
- The Vision Computer (TVCJ)
- IET Image Processing
- Journal of Electronic Imaging (JEI)

Workshop Reviewer

- IEEE International Workshop on Analysis and Modeling of Faces and Gestures Workshops (AMFG)

HONORS & AWARDS

- NSF I-Corps Grant, National Science Foundation 2022
- NVIDIA CCS Best Student Paper Award 2021
- 1st Place of the CVPR 2021 Challenge on Sign Language Recognition (both RGB & RGBD tracks) 2021
- 4th Rank in CVPR 2021 Challenge on Agriculture-Vision (supervised track) 2021
- PhD Network Grant, Northeastern University, USA 2019
- GapFund360 Award, Northeastern University, USA 2018
- NSF I-Corps Grant, National Science Foundation 2016
- JASSO Scholarship, Nagoya University, Japan 2014
- Outstanding Scholarship, Hong Kong Polytechnic University 2010, 2011, 2012, 2013

SKILLS

Languages:	English (full professional), Chinese (native), Cantonese, Japanese (basic).
Deep Learning Frameworks:	PyTorch (proficient), TensorFlow, CoreML (good knowledge).
Programming Languages:	Python, C/C++, C#, Java, HTML, JavaScript.
Others:	OpenCV, MATLAB, AWS E2 S3, Google Colab, Slurm, UWP, Git, etc.