

SHENG-WEI HUANG

San Francisco, CA; mlm4590027@gmail.com;

Education

The University of Texas at Dallas, Richardson, TX
Masters of Science in Computer Science

August 2019 – August 2021

National Tsing Hua University (NTHU), Hsinchu, Taiwan
Bachelor of Science in Electrical Engineering

September 2012 - June 2017

Professional Experience

Founder (Stealth AI Startup Exploration)

February 2026 – Present

- Explored an AI startup focused on applying agentic AI to physical routing workflows and EDA-native AI systems.
- Developed the technical vision and product strategy around integrating domain-specific workflows with agentic AI.

Applied Scientist, A9 Visual Search & AR

October 2021 – February 2026

Amazon Services Inc., 611 Cowper St, Palo Alto, CA 94301

- Researched and developed multimodal foundational models for e-commerce applications.
- Owned a production CLIP-based foundation model supporting Amazon Visual Search with business impact exceeding **\$2.5B**.
- Developed LLM- and VLM-based retrieval systems enabling multimodal query understanding and unified shopping retrieval.

Applied Scientist Intern, Visual Search & Augmented Reality Team

August 2020 – December 2020

Amazon Services Inc., 611 Cowper St, Palo Alto, CA 94301

Research Intern, Emergent Product Group

May 2020 – August 2020

Adobe Services Inc., 345 Park Avenue San Jose, CA 95110-2704

Algorithm Engineer, Advanced Creativity Team

July 2018 – July 2019

HTC Corporation, 88 Section 3, Zhongxing Road Xindian District, New Taipei City 231

Research Assistant, Computer Vision Laboratory

July 2017 – November 2017

National Tsing Hua University, 101, Section 2, Kuang-Fu Road, Hsinchu 300044

Selected Projects

Multimodal Large Language Model for Shopping Retrieval

January 2025 – February 2026

Core Search - Amazon Services Inc.

- Spearheaded development on MLLM-based representation learning solution for multimodal retrieval.
- Built scalable distributed training and evaluation infrastructure enabling efficient development of large-scale multimodal foundation models.
- Single-threaded owner of model development cycle from data collection, infrastructure setup to model training/evaluation and serving.

Agentic Multimodal Shopping Companion

June 2025 – August 2025

Core Search - Amazon Services Inc.

- Leveraged multimodal large language model for system orchestration and autonomous tool planning, enable tackling arbitrary tasks.
- Built autonomous tool planning and multimodal orchestration capabilities to support open-ended shopping workflows.
- Led proof-of-concept development demonstrating generalized task execution across multimodal shopping experiences.

Large Language Models (LLMs) for Multimodal Retrieval

January 2024 – November 2024

A9 Visual Search & AR, Amazon Services Inc.

- Leveraged LLMs for multimodal query understanding and retrieval system orchestration.
- Published in the FGVC11 workshop of **CVPR 2024**

Multimodal Foundational Models for Multimodal Retrieval

May 2023 – December 2023

A9 Visual Search & AR, Amazon Services Inc.

- CLIP-based architecture for multi-view e-commerce data to tackle street-to-shop retrieval problems.
- Productionized developed model and improved customer experiences multimodal search and image match.
- Published in **KDD 2024**: ACM SIGKDD Conference on Knowledge Discovery and Data Mining

Leveraging Generative Adversarial Networks for Data Augmentation

September 2017 – June 2018

Computer Vision Laboratory, NTHU (<https://bit.ly/2QkINld>)

- Developed novel generative adversarial network (GAN) for synthetic driving dataset generation.
- Published in the 15th European Conference on Computer Vision (**ECCV 2018**), Munich [1].

Publications

[1] Xinliang Zhu*, Sheng-Wei Huang*, Han Ding, Jinyu Yang, Kelvin Chen, Tao Zhou, Tal Neiman, Ouye Xie, Son Tran, Benjamin Yao, Douglas Gray, Anuj Bindal, and Arnab Dhua. 2024. Bringing Multimodality to Amazon Visual Search System. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24)

[2] Barbany, O., Huang, M., Zhu, X., & Dhua, A. (2024). Leveraging Large Language Models for Multimodal Search. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 1201-1210).

[3] Huang, S., Lin, C., Chen, S., Wu, Y., Hsu, P., & Lai, S., (2018), AugGAN: Cross Domain Adaptation with GAN-Based Data Augmentation. 15th European Conference on Computer Vision (ECCV), 2018

[4] H. J. Lin, S. Huang, S. Lai and C. Chiang, Indoor Scene Layout Estimation from a Single Image, 2018 24th International Conference on Pattern Recognition (ICPR)

[5] C. -T. Lin, S. -W. Huang, Y. -Y. Wu and S. -H. Lai, "GAN-Based Day-to-Night Image Style Transfer for Nighttime Vehicle Detection," in IEEE Transactions on Intelligent Transportation Systems, vol. 22, no. 2, pp. 951-963, Feb. 2021

Skills & Expertise

Languages: Python, C, C++, Java, Matlab

Tools: Pytorch, Tensorflow, Amazon Web Service (AWS)