

MAYA CHEN, Ph.D.

AI Research Scientist | Multimodal Learning | Trustworthy Generative AI

San Francisco, CA • +1 415 555 0176 • maya.chen@protonmail.com
linkedin.com/in/maya-chen-ai • github.com/mayachen-ai • Nexa Scholar

Research Profile

AI research scientist with 8+ years of experience advancing multimodal representation learning, efficient generative models, and reliable evaluation. Published 18 peer-reviewed papers at NeurIPS, ICML, ICLR, CVPR, and ACL (4,600+ citations; h-index 15) and transferred research into products serving 35M users. Combines rigorous experimentation with large-scale PyTorch training, human-centered evaluation, and technical leadership across research and engineering teams.

Research & Technical Expertise

Research	Multimodal learning; vision-language models; generative modeling; self-supervised learning; model evaluation; robustness; uncertainty; interpretability
Methods	Transformers; diffusion models; contrastive learning; RLHF/DPO; retrieval-augmented generation; distillation; parameter-efficient fine-tuning; causal inference
Programming	Python; PyTorch; JAX; NumPy; Hugging Face; CUDA/Triton; SQL; C++; Linux; Git
ML Systems	Distributed training (FSDP, DeepSpeed); Kubernetes; Ray; Spark; MLflow; Weights & Biases; Docker; AWS/GCP; reproducible experiment design

Research Experience

Staff Research Scientist <i>Luma Intelligence, Foundation Models Lab</i>	2022–Present <i>San Francisco, CA</i>
--	---

- Lead a six-person research program on controllable vision-language foundation models; developed a mixture-of-experts architecture that improved zero-shot visual reasoning by 11.8 points while reducing training FLOPs 27%.
- Created a preference-optimization method for image-grounded assistants using 420K human comparisons; reduced hallucinated object references 39% and increased expert-rated helpfulness 18% in a 2,000-prompt study.
- Co-designed a scalable data-curation pipeline for 1.8B image-text pairs using deduplication, semantic filtering, and privacy classifiers, increasing downstream sample efficiency 1.6×.
- Defined safety and robustness gates spanning calibration, demographic performance, prompt attacks, and memorization; the evaluation suite became a release requirement across four product teams.
- Published six papers, filed four patents, and shipped model improvements to experiences used by 35M monthly users; partnered with product and infrastructure teams from prototype through monitored launch.
- Mentor three research scientists and two interns; set quarterly research roadmaps and manage a 384-GPU experimental budget with 92% utilization.

Research Scientist <i>Aperture AI Research</i>	2019–2022 <i>Seattle, WA</i>
--	--

- Introduced CrossBind, a contrastive pretraining objective for fine-grained image-text alignment, improving retrieval Recall@1 by 14% and few-shot recognition by 8% across 12 benchmarks.
- Built uncertainty-aware evaluation for deployed document-understanding models; identified a silent domain-shift failure and enabled selective prediction that cut high-severity errors 46% at 94% coverage.
- Optimized mixed-precision distributed training for transformer models from 32 to 256 GPUs, raising throughput 2.3× and shortening a standard pretraining run from 19 days to 8 days.
- Authored five peer-reviewed papers (three first-author), open-sourced two libraries totaling 7,800 GitHub stars, and supported adoption by eight internal applied-ML teams.

Graduate Research Assistant <i>Stanford Vision and Learning Lab</i>	2015–2019 <i>Stanford, CA</i>
---	---

- Developed weakly supervised methods for learning compositional visual concepts from noisy web data, producing four publications and a 3.2M-example benchmark used by 40+ research groups.
- Designed controlled studies of shortcut learning and dataset bias; proposed counterfactual augmentation that improved worst-group accuracy by 13.5 points without group labels at inference.
- Collaborated with robotics researchers to transfer learned representations to manipulation, reducing labeled demonstrations required for novel-object grasping by 60%.

Education

Ph.D. in Computer Science

Stanford University

2019

Stanford, CA

Dissertation: *Learning Compositional Visual Representations from Weak Supervision*

Advisor: Prof. Elena Rossi

B.S. in Computer Science and Applied Mathematics, *summa cum laude*

University of Washington

2014

Seattle, WA

Selected Publications

- Chen, M.**, Adeyemi, T., Novak, J., and Rao, P. (2025). “Grounded Preference Optimization for Reliable Multimodal Assistants.” *Advances in Neural Information Processing Systems (NeurIPS)*. **Spotlight.**
- Chen, M.**, Singh, R., Morales, A., and Kim, S. (2024). “Sparse Experts for Compute-Efficient Vision–Language Pretraining.” *International Conference on Machine Learning (ICML)*.
- Okafor, N., **Chen, M.**, and Rao, P. (2024). “Calibrate Before You Deploy: Selective Generation under Multimodal Distribution Shift.” *International Conference on Learning Representations (ICLR)*. **Oral.**
- Chen, M.**, Alvarez, D., and Brooks, L. (2022). “CrossBind: Fine-Grained Contrastive Alignment beyond Global Image–Text Pairs.” *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Chen, M.**, Ibrahim, Y., and Rossi, E. (2019). “Counterfactual Augmentation for Robust Visual Recognition.” *Advances in Neural Information Processing Systems (NeurIPS)*.

Open-Source Research Software

EvalForge — evaluation toolkit for generative and multimodal models

github.com/mayachen-ai/evalforge | 5,200+ stars

Provides reproducible task suites, calibrated confidence metrics, bootstrap significance tests, and slice-based error analysis; adopted by 30+ academic and industry labs.

CrossBind — distributed contrastive-learning library

github.com/mayachen-ai/crossbind | 2,600+ stars

Implements memory-efficient multimodal pretraining, hard-negative mining, and streaming datasets with reference recipes from 8 to 256 GPUs.

Patents, Honors & Service

- 2025 Inventor, *Grounded Preference Learning for Multimodal Generation*, U.S. patent application filed
- 2024 Outstanding Paper Award, Workshop on Responsible Foundation Models, NeurIPS
- 2023 Luma Intelligence Research Impact Award for multimodal model reliability
- 2021 Rising Stars in AI, university research consortium honoree
- 2018 NSF Graduate Research Fellowship; Stanford Centennial Graduate Fellowship

Program Committee / Reviewer: NeurIPS, ICML, ICLR, CVPR, ACL, AAAI (2019–Present)

Area Chair: NeurIPS Datasets and Benchmarks Track (2024); CVPR (2025)

Community: Co-organizer, Bay Area Multimodal Learning Seminar; volunteer mentor, AI4ALL

Selected Invited Talks

- 2025 “Reliable Multimodal Models: From Preference Data to Deployment,” Berkeley AI Research
- 2024 “Scaling Vision–Language Learning with Sparse Experts,” MIT CSAIL
- 2023 “Evaluation as Infrastructure for Generative AI,” Women in Machine Learning Symposium, NeurIPS
- 2022 “Fine-Grained Alignment in Multimodal Representation Learning,” University of Toronto Vector Institute

Teaching & Mentorship

Research mentorship: Mentored 11 interns and junior researchers; six mentees became co-authors and four entered Ph.D. programs.

Guest lecturer: Stanford CS 231N, *Deep Learning for Computer Vision* (2021, 2023); developed lectures and exercises on contrastive multimodal learning.

Workshop instructor: Designed and taught a two-day course on trustworthy foundation-model evaluation for 120 research and engineering staff.