Jane Yang

San Diego, CA | j7yang@ucsd.edu | janeyang07.github.io | GitHub | Scholar

Ph.D. student who builds pipelines for large-scale multimodal data on naturalistic human behavior and uses human data to evaluate and benchmark AI models. My work spans 1,400+ hours of children's egocentric video, adult egocentric datasets, and synchronized eye-tracking, depth, IMU, and audio recordings, analyzed through human-model benchmarks, validation of model metrics against human judgments, representational analyses of model features, and controlled training experiments. Experienced in embodied perception, human attention, and model evaluation.

Education

Ph.D. in Experimental Psychology, UC San Diego. Advisor: Bria Long 2024–present

B.S. in Cognitive Science, UC San Diego. Machine Learning & Neural Computation; CS minor 2018–2022

Technical Skills

Programming	Python, R, C++, JavaScript/TypeScript
ML & Vision	PyTorch, HuggingFace; CLIP, DINOv3, YOLOE/YOLOv8, Qwen VLMs, MediaPipe, WhisperX
Evaluation & Modeling	Human-model benchmarking, validating model metrics against human judgments, VLM evaluation against gold sets, representational similarity analysis, embedding-space geometry and dispersion metrics, reinforcement learning models of behavior
Methods	Behavioral experiment design (web and in-lab), crowdsourced annotation, gold-set construction
Data & Infra	Multimodal data wrangling and synchronization (video, gaze, depth, IMU, audio), large-scale video processing (FFMPEG), SLURM GPU clusters, Docker, Git, automated testing
Sensing & Hardware	Head-mounted eye tracking (Pupil Labs Neon), depth cameras (Orbbec Femto Bolt), IMU, multi-sensor sync

Experience

Graduate Researcher, UC San Diego Visual Learning Lab 2024–present

- Built and maintain an end-to-end detection and embedding pipeline over 1,482 hours of egocentric video from 35 families: YOLOE detection, CLIP-based false-alarm filtering, crowdsourced validation, and CLIP/DINOv3 feature extraction on a GPU cluster; now replacing detection with a local Qwen VLM, benchmarked against a human-annotated gold set
- Co-led development of a CLIP-based vision-language alignment metric, validated it against human judgments, and applied it at scale to infant egocentric video, showing that well-aligned word-referent moments are rare relative to curated image-caption datasets (co-first-author preprint)
- Built and deployed MOCHI Kids, a human-model benchmark for 3D multiview object consistency adapted from the MOCHI shape-oddity task (TypeScript/jsPsych, Express/MongoDB, end-to-end tests)
- Showed that object statistics in children's visual input (frequency, exemplar variability, category geometry) diverge from adult egocentric video (VEDB) and curated benchmarks (ImageNet, LVIS, THINGS), using representational similarity analysis over 185 categories; first-author manuscript in revision at *Nature Human Behaviour*, with a public reproducibility capsule
- Separated viewpoint- from instance-driven within-category variability across 7,018 validated crops using global and local dispersion metrics; released as a reproducible package and interactive explorer (CCN 2026 spotlight); now designing controlled training experiments that manipulate each source to test effects on category learning in vision models
- Built a synchronized capture rig (two depth cameras, two eye trackers, room camera) for caregiver-child play to measure attention and object interaction; prototyping interaction measures from detections, hand tracking, depth, and IMU

Lab Technician, UT Austin Developing Intelligence Lab 2022–2024

- Built Python and MATLAB pipelines to process, analyze, and visualize multimodal streams from eye trackers, cameras, IMUs, and microphones recorded during naturalistic play in homes
- Fine-tuned YOLOv8 for egocentric object detection, deployed MediaPipe and OpenPose for hand and face tracking, and integrated WhisperX for transcription and diarization

Selected Publications

- **Yang, J.**, Sepuri, T., Tan, A.W.M., Aw, K.L., Frank, M.C., Long, B. (2026). Characterizing the visual representation of objects from the child's view. *In revision, Nature Human Behaviour.*
- Tan, A.W.M.*, **Yang, J.***, et al. (2025). Assessing the alignment between infants' visual and linguistic experience using multimodal language models. *Preprint*, arXiv:2511.18824.

Also: paper in *Developmental Science* (2025); co-organizer, *Beyond Curated Datasets* workshop, CCN 2026; invited talk, UNR (2026). Full list on Scholar.

Awards

NEI Early Career Scientist Travel Grant, VSS (2026); Anderson Travel Award (2025) and Fellowship (2024), UCSD