

Li Liu

Google Scholar | GitHub | Homepage: leolee7.github.io

Ph.D. Candidate in Computer Science and Engineering, UC Santa Cruz

Open to full-time roles starting 2027

lliu112@ucsc.edu

Multimodal LLMs · Agents · Spatial reasoning

RESEARCH PROJECTS

Spatial Intelligence (COLM 2026, NeurIPS 2024, one paper under review):

How multimodal models perceive space, build an internal map of it, and act when what they see is not enough.

- **Perceive Space:** Defined a new active-perception task, directional guidance, and built and open-sourced its 821-sample benchmark of real photos taken by blind users: VLMs answer even when the image cannot support an answer, instead of saying how to move the camera. Designed a label-free synthetic-data pipeline that works on any VQA dataset; fine-tuning on it raised open VLMs' accuracy by up to 16 percentage points, the best beating GPT-4o. [\[link\]](#)
- **Represent Space:** Built a benchmark that separates building a spatial layout from reasoning over it, and tested six LLMs: all read layouts better than they build them. Fine-tuning models to draw layouts more than doubled construction accuracy and improved text-only spatial reasoning, transferring to three external benchmarks; released the dataset. [\[link\]](#)
- **Ground a Real-Time Agent:** Built a real-time voice-and-touch agent on Gemini's native-audio model that grounds its answers in a map under the user's finger and runs on an everyday laptop; benchmarked its spatial grounding (83% correct on 393 spoken questions). Led the project from zero, from co-design with a blind collaborator to 2-hour in-person studies with 14 blind and low-vision adults and their qualitative analysis. Demoed at PitchAbility 2026. [\[link\]](#)

Uncertainty-Aware Prediction (IEEE TIM, EJRP Open): Proposed CPSC, a conformal predictor more accurate at lower compute than conformal k-NN, whose reliability scores select unlabeled data for augmentation [\[link\]](#); applied conformal prediction to COVID-19 ICU-risk triage that held up on new variants [\[link\]](#).

SELECTED PUBLICATIONS

- [1] **S. Huang***, **L. Liu*** et al. "Learning to Draw ASCII Improves Spatial Reasoning in Language Models." *COLM* 2026. (*co-first author) [\[link\]](#)
- [2] **L. Liu** et al. "Right This Way: Can VLMs Guide Us to See More to Answer Questions?" *NeurIPS* 2024. [\[link\]](#)
- [3] **L. Liu** et al. "TouchingSpace: An Audio-Haptic Map for Blind and Low-Vision Readers." *Under review; arXiv* 2026. [\[link\]](#)
- [4] **L. Liu** et al. "Pattern Discovery with Wide-Lens Analysis and Sharp-Focus Validation." *WWW Demo Track* 2026. [\[link\]](#)
- [5] **L. Liu** et al. "CPSC: Conformal Prediction With Shrunk Centroids for Efficient Prediction Reliability Quantification and Data Augmentation." *IEEE TIM* 2022. [\[link\]](#)

EXPERIENCE

Amazon

Seattle, WA

Applied Scientist Intern

June 2026 – Sept. 2026

- **Skills on Board: Maintaining Agent Skills as Typed Data (under review)** [\[link\]](#): Proposed and built a system that stores agent skills as typed, versioned data instead of free text, so people can trace and edit each step and agents follow the same source. Halved agent output tokens on SkillLearnBench at equal or better task performance.
- **Pipeline-Wide Failure Attribution:** Linked each skill to its evidence and rollouts so a failure can be traced to the agent, the task, or the benchmark; this found defects in SkillLearnBench itself.

Amazon

Santa Clara, CA

Applied Scientist Intern

June 2025 – Sept. 2025

- **Taxonomy Discovery from Unlabeled Conversations (WWW 2026 Demo)** [\[link\]](#): Designed a modular, human-in-the-loop framework that evolves its own taxonomy from a few seed keywords, alternating a wide-lens view of where patterns cluster with sharp-focus checks of each extraction; every pattern traces to its source.

Amazon

Palo Alto, CA

Applied Scientist Intern

June 2024 – Sept. 2024

- **Evidence-Grounded Search from Reviews** [\[link\]](#): Built a search system that turns customer reviews into search facets, each linked to the review sentence behind it, to answer needs catalogs do not record, such as a dish kids will eat. Later shown on recipes to find 18% more matches than keyword search.

Nissan Research Center

Santa Clara, CA

Data Scientist Intern

July 2023 – Sept. 2023

- **Driver Engagement in Autonomous Driving:** Developed a time-series model of driver engagement from real-world driving logs, improving prediction accuracy by 16 percentage points.

EDUCATION

UC Santa Cruz, Ph.D. in Computer Science and Engineering, advisor: Leilani Gilpin	2022 – 2027 (est.)
Zhejiang University, M.Eng. in Control Science and Engineering	2019 – 2022
Zhejiang University, B.S., Chu Kochen Honors College	2015 – 2019

ACADEMIC ACTIVITIES

Reviewer: NeurIPS, ICLR, CVPR, ECCV, CIKM, WACV. **TA:** Python, Computer Vision. **Mentor:** Foothill SLI, UCSC SIP, Stanford summer research.