

MANASI SHARMA

Researcher at Scale AI; MS in CS from Stanford University; BA in CS from Columbia University
Generative AI, Agents, Natural Language Processing, Reinforcement Learning, Machine Learning, Computer Vision
manasis@cs.stanford.edu | www.linkedin.com/in/manasi1 | manasi-sharma.github.io | San Francisco, CA

EDUCATION

Stanford University, California M.S. in Computer Science (AI/ML Track), GPA: 3.96/4.00 Research: <i>Language-Conditioned Diffusion Models for Robot Learning</i> under Prof. Dorsa Sadigh	Sep '21 - Jun '23
Columbia University, New York B.A. in Computer Science with concentration in Physics, GPA: 3.81/4.00	Aug '17 - Jun '21

WORK EXPERIENCE

Scale AI <i>ML Research Engineer, Reasoning & Agents Research Team, Scale Labs</i> <ul style="list-style-type: none">Researched on computer-use & browser-use agents, spearheading 3 first-author publications (7 total) on evaluations at top-tier conferences (ICLR, ICML '26); including ResearchRubrics, referenced in a Perplexity AI release.Collaborated with external organizations like the Center for AI Safety at Berkeley on benchmarks for economically useful tasks (Remote Labor Index) and tool-use (MCP-Atlas), which was referenced in OpenAI & Anthropic's latest model cards.Post-trained models with RL to analyze strategies for optimal data learning and contributed to research on an efficient RL algorithm with intermediate rewards, leading to a publication at a NeurIPS '25 workshop.Collaborated with researchers at top frontier labs, offering consultation and guidance on advanced computer-use data types involving programmatic & agentic verifiers.	Mar '25 - Present San Francisco, CA
MIT Lincoln Laboratory <i>Research Scientist, Artificial Intelligence Technology Group</i> <ul style="list-style-type: none">Spearheaded a project on LLM explainability & trust that was published at a ICML'24 workshop, & built an open-source LLM-testing framework (maite). Participated in a project collaboration with MIT Profs. Philip Isola & Jacob Andreas.Architected an internal tool for efficient evaluation of a Retrieval Augmented Generation (RAG) system with 15+ distinct metrics & assessed it on referencing past recordings in a speech-to-text application, achieving 92% accuracy.	Aug '23 - Mar '25 San Francisco, CA (Remote)
Nissan Research <i>Research Intern, Autonomous Systems, Alliance Innovation Lab</i> <ul style="list-style-type: none">Engineered an end-to-end LiDAR 3D point-cloud classification system in Python & C++ for Nissan Autonomous Vehicles, which achieved >95% accuracy, ~2% FPR and 85% reduction in runtime on classifying real-world cars, pedestrians, cyclists, etc. The system has been deployed in Nissan Autonomous Vehicles beginning Winter '22.	Jun '22 - Sep '22 Santa Clara, CA
Google <i>CS Research Mentorship Program</i> <ul style="list-style-type: none">Ideated research projects on language models & diffusion models as part of the CS Research Mentorship program.	Mar '22 - Jun '22 Mountain View, CA

PUBLICATIONS

• M. Sharma , C.B. Zhang, C. Bandi, et. al, "ResearchRubrics: A Benchmark of Prompts and Rubrics For Evaluating Deep Research Agents". ICLR 2026 Main Conference Poster .	Nov '25
• M. Sharma , B. Kenstler, B. Liu, "Rubric Robustness: Evaluating the Sensitivity of Rubrics-Based Benchmarks. to Simple Perturbations". ICML 2026 Main Conference Poster .	Mar '26
• M. Sharma , D. Zhang, B. Liu, "Agentic Rubrics for Test-Time Scaling of Computer-Use Agents". <i>Submission to NeurIPS in progress.</i>	May '26
• M. Mazeika, A. Gatti, ..., M. Sharma , et. al, "Remote Labor Index: Measuring AI Automation of Remote Work". Preprint .	Oct '25
• V. Nath, E. Lau, A. Gunjal, M. Sharma , et. al, "Adaptive Guidance Accelerates Reinforcement Learning of Reasoning Models". NeurIPS 2025 1st Workshop on Efficient Reasoning .	Jun '25
• C. Bandi, B. Hertzberg, ..., M. Sharma , et. al, "MCP-Atlas: A Large-Scale Benchmark for Tool-Use Competency with Real MCP Servers". Submission to NeurIPS in progress.	Jan '26

- S. Kundurthy, C. Na, ..., **M. Sharma**, et. al, "SpreadsheetArena: Decomposing Preference in LLM Generation of Spreadsheet Workbooks". [ICML 2026 Main Conference Poster](#). Mar '26
- **M. Sharma**, HC Siu, et. al. "Why Would You Suggest That? Human Trust in Language Model Responses". [ICML 2024 Humans, Algorithmic Decision-Making and Society Workshop](#). Jun '24
- **M. Sharma**, Y. Cui, "Exploring and Improving the Spatial Reasoning Abilities of Large Language Models". [NeurIPS 2023 Instruction Tuning and Following Workshop](#). Oct '23
- C. Li, C. Gokmen, ..., **M. Sharma**, et. al, "BEHAVIOR-1K: A Benchmark for Embodied AI with 1,000 Everyday Activities and Realistic Simulation". [Conference on Robot Learning \(CoRL\) 2022](#). Nominated for 'Best Paper'. Jun '22
- J. Matilla, **M. Sharma**, D. Hsu, Z. Haiman, "Interpreting deep learning models for weak lensing". [Physical Review D](#), American Physical Society, 102(12). Dec '20

RESEARCH EXPERIENCE

Stanford University, Stanford Vision Laboratory & ILIAD Robotics Lab Sep '21 - Jun '23
Research Intern (Prof. Dorsa Sadigh, Prof. Fei-Fei Li, Prof. Jiajun Wu)

- Co-led a project on using few-shot prompting for LLMs to recognize patterns in trajectory data (NIPS workshop '23) & another on using diffusion models for trajectory generation conditioned on Language.
- Led the development of the Knowledgebase for [iGibson](#) and [BEHAVIOR-1K](#), an ImageNet-scale robotic simulation benchmark. Accepted for CoRL '22 and nominated for 'Best Paper' award. Presented live tutorial at ECCV '22.

Columbia University, Data Science Institute Sep '19 - Jun '21
Research Intern, Dept. of CS & Astronomy (Prof. Daniel Hsu and Prof. Zoltan Haiman)

California Institute of Technology, Division of Physics, Mathematics and Astronomy Jun '19 - Aug '19
Visiting Undergraduate Research Program (VURP) Intern, Palomar Gattini-IR Group (Prof. Mansi Kasliwal)

TECHNICAL SKILLS

- Programming Languages: Proficient: Python, C++/C, ROS, CUDA, Java, JavaScript, LaTeX; Familiar: Julia, SQL, SQLite
- Frameworks: TensorFlow, Keras, PyTorch, Scikit-Learn, NLTK, PyBullet, MeshLab, NetworkX, PyG, OpenCV
- Tools: Colab/GCP, Jupyter Notebooks, Visual Studio, Git, MySQL (Familiar), Figma

HONORS

- 1 of 25 students accepted to the highly selective Pear Garage program for Entrepreneurship at Pear VC Oct '22
- 1 of 18 accepted to the highly selective [GFSD](#) (Graduate Fellowships for STEM Diversity) Program Mar '22
- 1 of 50 accepted into Google's CS Research Mentorship Program ([CSRMP](#)), Class of 2022A Feb '22
- Selected for the final round of the GEM Fellowship Jan '22
- Dean's List (in 6 out of 7 graded semesters, awarded to top 20%), Columbia University Fall '17 - Fall '20
- Columbia Undergraduate Research Fellowship (URF), Columbia College Summer Funding Program May '20
- Visiting Undergraduate Research Program (VURP) Award, California Institute of Technology May '19
- 1 of 25 awarded Laidlaw Undergraduate Research & Leadership Scholarship, Columbia Univ. '18 - '19
- Andy Grove Scholarship for Intel Employees' Children, Intel Foundation Fall '19

OTHER

Teaching Experience

- **Stanford University, School of Engineering, Stanford, CA** Mar '22 - Jun '23
Graduate Teaching Assistant for over 4 quarters for 3 of the most popular CS classes at Stanford (>600 students):
 - [C224N](#) (NLP, Prof. Manning), [CS231N](#) (CV, Prof. Fei-Fei Li) and [CS230](#) (Deep Learning, Prof. Andrew Ng)
- **Columbia University, Department of Mathematics, New York, NY** Sep '19 - Jun '21
Undergraduate Teaching Assistant for Calculus III (across 4 semesters)

Languages

- Hindi (fluent), Spanish (intermediate)