

SACHIT MENON

sachit.menon@columbia.edu | sachit-menon.github.io

Researcher in multimodal foundation models, LLM agents, and inference-time scaling. Focused on improving reasoning, tool use, and trust in large pretrained systems. Published at CVPR, ICCV (Oral), ICLR (Oral), EMNLP with 4000+ citations; contributor to Gemini 2.5.

EDUCATION

Columbia University, New York City, New York (GPA: 4.02/4.00)

Ph.D., Computer Science, Ongoing (Spring 2026 Graduation)

M.S., Computer Science, May 2022

- Conducting research in large pretrained models; advised by Prof. Carl Vondrick, collaboration with Profs. Rich Zemel, David Blei, Micah Goldblum

Duke University, Durham, North Carolina (GPA: 3.97/4.00)

B.S., Mathematics and Computer Science, May 2020

Graduation with Highest Distinction; Summa Cum Laude

RESEARCH EXPERIENCE

CVLab (Columbia University), Deep Learning Researcher

July 2020 – Present

New York, New York

- Primary focus: multimodal foundation models and novel inference methods
- Recent work explores reinforcement learning for LLM agent tool-calling, safety, inference-time compute scaling and guardrails
- Applying large language models to vision tasks, e.g., with code generation
- Featured #1 on HackerNews, >1.7k GitHub stars

Google DeepMind, Student Researcher

Sep. 2024 – March 2025

Cambridge, MA

- Pushed frontier research to measure and improve video reasoning capabilities, particularly with agentic harnesses and tool use using multimodal large language models; mentored by Arsha Nagrani, Ahmet Iscen, and Cordelia Schmidt
- Contributed to Gemini 2.5 (multimodal post-training + evaluations)
- Extended OneTwo, GDM's open-source agent framework, to visual agents
- Published Minerva video reasoning benchmark at ICCV 2025; second preprint in submission (CAViAR)

Meta Generative AI, Research Scientist Intern

May 2023 – Dec. 2023

New York, New York

- Trained diffusion + language models to generate mixed text and images for instructional articles mentored by Rohit Girdhar and Ishan Misra
- Published Generating Illustrated Instructions at CVPR 2024
- Performed finetuning, inference, and evaluation for diffusion models + LLMs in distributed GPU multi-node settings

Rudin Lab (Duke University), Deep Learning Research Asst

January 2018 – May 2020

Durham, North Carolina

- Transfer learning with CNNs, deep generative models for image super-resolution
- Released code with >6.7k GitHub stars, top 5 posts of all time on Reddit r/MachineLearning after first week of release (redd.it/hciw10)

AWARDS

NSF Graduate Research Fellow | CAIRFI PhD Fellow | Columbia Presidential Fellow

Angier B. Duke Scholar: Duke's flagship merit scholarship awarded to ~15/32,000 undergraduates/year | **Alex Vasilos Award:** Highest award given by Duke CS

Barry M. Goldwater Scholar | CRA Outstanding Undergraduate Research Award (Finalist)
Rhodes, Marshall Scholarship Finalist | Hertz Fellowship Semifinalist (2019)

SELECTED PUBLICATIONS

Gemini 2.5: Pushing the Frontier With Advanced Reasoning, Multimodality, Long Context, and Next Generation Agentic Capabilities. Technical Report. Gemini Team. <https://arxiv.org/abs/2507.06261>

Minerva: Evaluating Complex Video Reasoning. ICCV 2025. Arsha Nagrani, **Sachit Menon**, Ahmet Iscen, Shyamal Buch, Ramin Mehran, Nilpa Jha, Anja Hauth, Yukun Zhu, Carl Vondrick, Mikhail Sirotenko, Cordelia Schmid, Tobias Weyand. <https://arxiv.org/abs/2505.00681>

CAViAR: Critic-Augmented Video Agentic Reasoning. In Submission. **Sachit Menon**, Ahmet Iscen, Arsha Nagrani, Tobias Weyand, Carl Vondrick, Cordelia Schmid. <https://arxiv.org/abs/2509.07680>

Whiteboard-of-Thought: Thinking Step-by-Step Across Modalities. EMNLP 2024. **Sachit Menon**, Richard Zemel, Carl Vondrick. <https://arxiv.org/abs/2406.1456>

Generating Illustrated Instructions. CVPR 2024. **Sachit Menon**, Ishan Misra, Rohit Girdhar. <https://arxiv.org/abs/2312.04552>

ViperGPT: Visual Inference via Program Execution for Reasoning. ICCV 2023 (**Oral**). **Sachit Menon***, Didac Suris*, Carl Vondrick. <https://arxiv.org/abs/2210.07183>

Doubly Right Object Recognition: A Why Prompt for Visual Rationales. CVPR 2023. Chengzhi Mao, Revant Teotia, Amrutha Sundar, **Sachit Menon**, Junfeng Yang, Xin Wang, Carl Vondrick. <https://arxiv.org/abs/2212.06202>

Shadows Shed Light on 3D Objects. CVPR 2023. Ruoshi Liu, **Sachit Menon**, Carl Vondrick. <https://arxiv.org/abs/2206.08990>

Visual Classification via Description from Large Language Models. ICLR 2023 (**Notable Top 5% - Oral**). **Sachit Menon**, Carl Vondrick. <https://arxiv.org/abs/2210.07183>

Task Bias in Vision-Language Models. IJCV 2023. **Sachit Menon***, Ishaan Chandratreya*, Carl Vondrick. <https://arxiv.org/abs/2212.04412>

Affective Faces for Goal-Driven Dyadic Communication. In Submission. Scott Geng, Revant Teotia, Purva Tendulkar, **Sachit Menon**, Carl Vondrick

Contrastive Learning Effectively Mitigates Posterior Collapse. UAI 2022. **Sachit Menon**, David M. Blei, Carl Vondrick. <https://arxiv.org/abs/2207.09535>

PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models. CVPR 2020. **Sachit Menon***, Alex Damian*, Nikhil Ravi, Shijia Hu, Cynthia Rudin. <https://arxiv.org/abs/2003.03808>

Tools: Python, PyTorch, HuggingFace Transformers/Accelerate, vLLM, TensorFlow/Keras, Numpy, Scipy, Scikit-Learn, Matplotlib, Seaborn, Pandas