Sanchit Sinha
ss7mu@virginia.edu | github.com/sanchit97 | sanchitsinha.com |
scholar.google.com/citations?user=squ4_6IAAAAJ&hl=en

EDUCATION

UNIVERSITY OF VIRGINIA Charlottesville, Virginia
Doctor of Philosophy (Ph.D.) *in Computer Science* 05/2021 - 05/2026 (expected)
Advised by Dr. Aidong Zhang - improving interpretability, explainability, concept extraction for VLMs and structured data
Master of Science (M.S) in Computer Science GPA: 4.0/4.0 08/2019 - 05/2021
IIIT-DELHI New Delhi, India
Bachelor of Technology (B. Tech.) in Computer Science with Honors 08/2015 - 05/2019

OPEN SOURCE PROJECTS

ChartRVR - Post-training Multimodal LLMs for Chart Understanding (**800+** downloads on HuggingFace)
Work done as part of Morgan Stanley Research group `https://github.com/sanchit97/chartrl`. First approach to post-train VLMs on structured data using GRPO with verifiable rewards. `https://huggingface.co/sanchit97/chart-rvr-3b`

WORK EXPERIENCE

Morgan Stanley New York City, NY, USA
Machine Learning Research Intern 06/2025 – 08/2025
- Improving structured data understanding in multimodal LLMs using reinforcement learning with verifiable rewards (GRPO)
- Beating direct, chain-of-thought prompting by 10%+, with structured data tailored reward design and data curation
- Seminal work on RLVR design and implementation on efficient models esp. financial charts - candlestick, time series, etc.

Amazon AGI Cambridge, MA, USA
Applied Scientist Intern 05/2023 – 08/2023
- Improving warmup approaches for improved in-context learning performance using second-order meta-learning approaches
- Beating standard meta-training approaches by a baseline minimum of 3%, a challenging feat not discussed before
- Seminal work on exploring dual optimization landscape in LLMs. Formalized insights on task selection, optimization, etc.

Amazon Web Services (AWS), Amazon Sunnyvale, CA, USA
Applied Scientist Intern, AWS Lex 05/2022 – 08/2022
- Implemented parameter efficient self-supervised accent domain adaptation on large speech models (HuBERT) using adapters
- Demonstrated improved performance on downstream speech tasks using general fine-tuning data by minimum 5%
- Improved generic accent information learned by large speech models without explicit labeling - reducing manual annotation

Unity Technologies (Unity 3D) Seattle, WA, USA
ML-Computer Vision Intern, AI@Unity 05/2020 – 08/2020
- Implemented a real-time video object tracking segmentation model (multimodal) with benchmark performance
- Containerized deployment on GCP/AWS with ETL functionality, robust fine-tuning, and scalable pipelining (Kubeflow)
- Designed multi-domain (including synthetic data) training algorithms (domain randomization) for better generalizability

PUBLICATIONS - Best viewed in Google Scholar

- **Multimodal LLMs (VLMs), Explainability and Alignment:**
  - **Sinha, Sanchit**, Xiong, G. and Zhang, A. “Concept-RuleNet: Grounded Multi-Agent Neurosymbolic Reasoning in Vision Language Models”, AAAI Conference on Artificial Intelligence 2026 (**AAAI ’26 (oral)**)
  - **Sinha, Sanchit**, Xiong, G. and Zhang, A.“COCO-Tree: Compositional Hierarchical Concept Trees for Enhanced Reasoning in Vision-Language Models” Empirical Methods in Natural Language Processing (**EMNLP ’25 (main)**)
  - **Sinha, Sanchit**, Xiong, G. and Zhang, A.“ASCENT-ViT: Attention-based Scale-aware Concept Learning Framework for Enhanced Alignment in Vision Transformers”, International Joint Conferences on AI **(IJCAI ’25)**
  - He, Zhenghao, **Sinha, Sanchit**, Xiong, G. and Zhang, A. “GCAV: A Global Concept Activation Vector Framework for Cross-Layer Consistency Interpretability” International Conference on Computer Vision. 2025 **(ICCV ’25)**
  - **Sinha, Sanchit**, Xiong, G. and Zhang, A. “CoLiDR: Concept Learning using Aggregated Disentangled Representations.” 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024 **(KDD ’24)**.
- **LLM and M-LLM/VLM Finetuning (Post-training):**
  - **Sinha, Sanchit** et al. “Understanding and Improving Chain-of-Thought Reasoning Dynamics in Large Vision Language Models” (Under Review at ACL ’26)
  - **Sinha, Sanchit** et al. “Chart-RVR: Reinforcement Learning with Verifiable Rewards for Explainable Chart Reasoning” (Arxiv preprint / Under Review at ICLR ’26)
  - **Sinha, Sanchit**, et al. “MAML-en-LLM: Model Agnostic Meta-training of LLMs for Improved In-context Learning.” Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024 **(KDD ’24)**.
  - Bhatia, Anshu*., **Sinha, Sanchit***., et al. “Don’t stop self-supervision: Accent adaptation of speech representations via residual adapters.” **(Interspeech ’23)**.

  - Sun, Jianhui, **Sinha, Sanchit** and Zhang, A. “Enhance Diffusion to Improve Robust Generalization.” Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023 **(KDD ’23)**.
- **Robustness/Domain Shifts:**
  - Xiong G., He Z., Liu B., **Sinha, Sanchit** and Zhang A. “Toward Faithful Retrieval-Augmented Generation with Sparse Autoencoders” (Arxiv preprint / Under Review at ICLR ’26)
  - Xiong, Guangzhi, **Sinha, Sanchit**, and Zhang, Aidong., Neural Additive Experts: Context-Gated Experts for Controllable Model Additivity (**AISTATS ’26**)
  - **Sinha, Sanchit**, Xiong, G. and Zhang, A. 2024. “A Self-Explaining Neural Architecture for Generalizable Concept Learning.” Thirty-Third International Joint Conference on Artificial Intelligence **(IJCAI ’24)**.
  - **Sinha, Sanchit**, et al. “Understanding and enhancing robustness of concept-based models.” AAAI Conference on Artificial Intelligence, 2023 **(AAAI ’23)**.
  - Xiong, Guangzhi, **Sinha, Sanchit**, and Zhang, Aidong. “ProtoNAM: Prototypical Neural Additive Models for Interpretable Deep Tabular Learning.” **(TKDD Journal)**
  - **Sinha, Sanchit**, et al. ”Perturbing Inputs for Fragile Interpretations in Deep Natural Language Processing.” Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP. 2021. **(EMNLP-Blackbox ’21)**
  - Agarwal*, M., **Sinha*, S.**, et al. “Triplet transform learning for automated primate face recognition.” **(ICIP ’19)**
  - **Sinha, Sanchit**, et al. “Exploring bias in primate face detection and recognition.” **(ECCV-W ’19)**

PRE-PRINTS

- **Sinha, Sanchit**, Guangzhi Xiong, and Aidong Zhang. “Structural Causality-based Generalizable Concept Discovery Models” (Arxiv): Utilizing Structural Causal Frameworks to improve the interpretability of Variational Autoencoders.
- **Unsupervised Image to Image Translation using GANs**
  Add semi-supervision in unsupervised (CycleGAN) to obtain a super-linear increase in performance with respect to supervised methods

AWARDS

Student Travel Award - KDD 2024, AAAI 2023. (< 20% selection rate)
Amazon Conference Grant - 2024
Cohere Project Grant $1000 - 2024
Reviewer - NeurIPS, ICML, ICLR, CVPR, ECCV, ICCV, KDD, EMNLP (multiple times, 2022-present)
School of Engineering and Applied Science - PhD Fellowship 2021-22

Other OSS work

ChartRL - Repository for Post-training VLMs for Chart Understanding
Work done as part of Morgan Stanley Research group `https://github.com/sanchit97/chartrl`. First approach to post-train VLMs and MLLMs on structured data. Implementation of GRPO (and DPO) style training.

FFmpeg - Google Summer of Code, 2017 Remote
Student Developer 05/2017 – 08/2017

- Nominated in a highly selective student open source developer program hosted by Google (code on Github profile)
- Designed/implemented audio processing decoder for Ambisonic AR-sound files to custom speaker array configuration