

Sanchit Sinha

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

EDUCATION

UNIVERSITY OF VIRGINIA

Doctor of Philosophy (Ph.D.) in Computer Science

Charlottesville, Virginia

05/2021 - 01/2026 (expected)

Advised by Dr. Aidong Zhang - improving interpretability, explainability, adversarial robustness, and concept extraction.

Master of Science (M.S) in Computer Science

GPA: 4.0/4.0

08/2019 - 05/2021

Elective Courses: Advanced Deep Learning, Machine Learning, Data Mining, NLP, Manifold Analysis, Graph Mining

IIIT-DELHI

New Delhi, India

Bachelor of Technology (B. Tech.) in Computer Science with Honors

GPA: 8.28/10

08/2015 - 05/2019

Elective Courses: Advanced ML, Artificial Intelligence, Parallel Programming, Advanced Algos, Collab Filtering, Biometrics

PUBLICATIONS - Best viewed in Google Scholar

- **Sinha, Sanchit**, Xiong, G. and Zhang, A. "ASCENT-ViT: Attention-based Scale-aware Concept Learning Framework for Enhanced Alignment in Vision Transformers" In Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence (**IJCAI '25**)
- **Sinha, Sanchit**, Xiong, G. and Zhang, A. "CoLiDR: Concept Learning using Aggregated Disentangled Representations." Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. 2024 (**KDD '24**).
- **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**).
- **Sinha, Sanchit**, Xiong, G. and Zhang, A. 2024. "A Self-Explaining Neural Architecture for Generalizable Concept Learning." In Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (**IJCAI '24**).
- **Sinha, Sanchit**, et al. "Understanding and enhancing robustness of concept-based models." Proceedings of the AAAI Conference on Artificial Intelligence, 2023 (**AAAI '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**).
- Bhatia, Anshu*, **Sinha, Sanchit***, Dingliwal, S., Gopalakrishnan, K., Bodapati, S., and Kirchhoff, K. (2023). "Don't stop self-supervision: Accent adaptation of speech representations via residual adapters." Proceedings of Interspeech, 2023. (**Interspeech '23**).
- **Sinha, Sanchit**, et al. "Perturbing Inputs for Fragile Interpretations in Deep Natural Language Processing." Proceedings of the Fourth BlackboxNLP Workshop on Analyzing and Interpreting Neural Networks for NLP. 2021. (**EMNLP-Blackbox '21**)
- Agarwal*, M., **Sinha***, S., Singh, M., Nagpal, S., Singh, R., and Vatsa, M. "Triplet transform learning for automated primate face recognition." In 2019 IEEE International Conference on Image Processing (ICIP). (**ICIP '19**)
- **Sinha, Sanchit**, et al. "Exploring bias in primate face detection and recognition." Proceedings of the European Conference on Computer Vision (ECCV) Workshops. 2018. (**ECCV-W '19**)

WORK EXPERIENCE

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)

ML-Computer Vision Intern, AI@Unity

Seattle, WA, USA

05/2020 – 08/2020

- Implemented a real time video object tracking segmentation model with benchmark performance on public leaderboards
- 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

FFmpeg - Google Summer of Code, 2017

Student Developer

Remote

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

PRE-PRINTS/UNDER REVIEW

- **Sinha, Sanchit**, Xiong, Guangzhi, He, Zhenghao, Zhang, Aidong Symbol-VLM: Visual Concept-Conditioned Symbol Generation for Enhanced Neuro-Symbolic Reasoning (Under Review)
- Xiong, Guangzhi, **Sinha, Sanchit**, and Zhang, Aidong. “ProtoNAM: Prototypical Neural Additive Models for Interpretable Deep Tabular Learning.” arXiv preprint arXiv:2410.04723 (2024). (Under Review - TKDD)
- **Sinha, Sanchit**, Guangzhi Xiong, and Aidong Zhang. “Structural Causality-based Generalizable Concept Discovery Models” (Arxiv)

AWARDS

Student Travel Award - KDD 2024, AAAI 2023. (20% selection rate)

Amazon Conference Grant - 2024

Cohere Project Grant \$1000 - 2024

Reviewer - NeurIPS, ICML, ICLR, KDD, EMNLP (2022-present)

School of Engineering and Applied Science - PhD Fellowship 2021-22

ONGOING PROJECT WORK

Neurosymbolic Concept-based Reasoning with LLMs *Under review, ICML '25*

Using LLM-Agents to extract, ground, and compose concepts into neurosymbolic entities for better explainability and predictions in low-resource VLMs. Seminal work linking grounding of concepts and neuro-symbolic reasoning.

Advancing Additive Models with Mixture of Experts (MoEs) *Under review, KDD '25*

Utilizing a Mixture of Experts as a tool to combine additive model features and model interactions improving performance

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