

# Nihal Jain

✉ [nihal.jain@columbia.edu](mailto:nihal.jain@columbia.edu) • 🌐 [nihaljn.github.io](https://nihaljn.github.io)

## Education

---

### Columbia University

*Ph.D. in Computer Science, GPA: 4.0/4.0*

**New York, NY**

*Sep 2025 – Present*

### Carnegie Mellon University

*M.S. in Machine Learning, GPA: 4.08/4.33*

**Pittsburgh, PA**

*Aug 2021 – Dec 2022*

### Birla Institute of Technology & Science, Pilani – Hyderabad Campus

*B.E. in Computer Science, GPA: 9.94/10.0*

**Hyderabad, India**

*Aug 2017 – Jul 2021*

## Selected Publications

(\* = equal contribution)

### On Mitigating Code LLM Hallucinations with API Documentation

*Nihal Jain, Robert Kwiatkowski, Baishakhi Ray, Murali Krishna Ramanathan, Varun Kumar*

International Conference on Software Engineering (ICSE) 2025 – Software Engineering in Practice Track

### The Amazon Nova family of models: Technical report and model card

*Aaron Langford, ..., Nihal Jain, ... (many authors)*

Amazon Technical Reports 2024

### CrossCodeEval: A Diverse and Multilingual Benchmark for Cross-File Code Completion

*Yangruibo Ding\*, Zijian Wang\*, Wasi Ahmad\*, Hantian Ding, Ming Tan, Nihal Jain, Murali Krishna Ramanathan, Ramesh Nallapati, Parminder Bhatia, Dan Roth, Bing Xiang*

NeurIPS 2023 – Datasets and Benchmarks Track

### ContraCLM: Contrastive Learning For Causal Language Model

*Nihal Jain\*, Dejiao Zhang\*, Wasi Uddin Ahmad\*, Zijian Wang, Feng Nan, Xiaopeng Li, Ming Tan, Ramesh Nallapati, Baishakhi Ray, Parminder Bhatia, Xiaofei Ma, Bing Xiang*

Association for Computational Linguistics (ACL) 2023

### Multiviz: Towards visualizing and understanding multimodal models

*Paul Pu Liang, Yiwei Lyu, Gunjan Chhablani, Nihal Jain, Zihao Deng, Xingbo Wang, Louis-Philippe Morency, Ruslan Salakhutdinov*

International Conference on Learning Representations (ICLR) 2023

### Self-supervised Multi-view Disentanglement for Expansion of Visual Collections

*Nihal Jain, Praneetha Vaddamanu, Paridhi Maheshwari, Vishwa Vinay, Kuldeep Kulkarni*

Web Search and Data Mining (WSDM) 2023

## Experience

---

### Amazon Web Services (AWS)

New York, NY

Applied Scientist II

Oct 2024 – Aug 2025

Applied Scientist I

Feb 2023 – Sep 2024

- Led post-training of LLMs for code generation in the Amazon Q Developer/CodeWhisperer team
- Designed experiments to improve LLMs for code editing and software engineering agentic tasks
- Published research on code generation with LLMs at venues like NeurIPS 2023, ICSE 2025

Applied Scientist Intern

May 2022 – Aug 2022

- Improved generative language models on discriminative tasks using contrastive learning (ACL 2023)

### Adobe Research

Bangalore, India (Remote)

Research Intern

Jan 2021 – Jul 2021

- Self-supervised representation learning for multi-modal information retrieval (WSDM 2023)

Research Intern

May 2020 – Aug 2020

- Developed GANs to generate color profiles condition on textual queries for image search (CIKM 2021)

### Indian Space Research Organisation (ISRO)

Dehradun, India

Research Intern

May 2019 – Jul 2019

- Implemented semantic segmentation of satellite images of Indian terrain (Geocarto Intl. 2020)

## Awards

---

### Institute Gold Medal – BITS, Pilani

Aug 2021

- Awarded for obtaining the 1<sup>st</sup> position in the class of ~1100 graduating students

### KC Mahindra Scholarship for Post Graduate Studies Abroad

Jul 2021

- Granted to pursue post-graduate studies abroad offered to ~50 out of >1000 applicants

## Invited Talks

---

### Code Generation Hallucinations & Alignment: An Industry Perspective 🌐

Nov 2024

- Columbia University, Host: Prof. Baishakhi Ray

## Service

---

### Program Committee

2023 – Present

- NeurIPS (2023, 2025), ARR (Feb 2024, Jun 2024), ICML (2024, 2025), AAAI (2025), ICLR (2025)

### Admissions Committee – CMU Machine Learning Department

2021 – 2022

- Reviewed applications to the M.S. in Machine Learning program for the Fall 2022 intake