

Naman Jain

Website: naman-ntc.github.io ◊ Email: naman_jain@berkeley.edu

EDUCATION

University of California, Berkeley <i>Ph.D. in Computer Science</i>	<i>2022 - present</i>
Indian Institute of Technology, Bombay <i>B.Tech (Honors) in Computer Science and Engineering</i>	<i>2016 - 2020</i>

WORK AND INTERNSHIPS

AWS CodeWhisperer, Sunnyvale <i>Applied Research Intern</i>	<i>June '23 - Aug '23</i>
Microsoft Research (MSR), India <i>Research Fellow</i>	<i>Aug '20 - Aug '22</i>
University of Maryland, College Park <i>Research Intern</i>	<i>June '19 - Aug '19</i>

PUBLICATIONS

8. How to teach LLMs new knowledge: finetune LLMs to read or to memorize?
Tianjun Zhang, Shishir G. Patil, **Naman Jain**, Sheng Shen, Matei Zaharia, Ion Stoica and Joseph E. Gonzalez
Preprint available on request
7. LLM-Assisted Code Cleaning For Training Accurate Code Generators
Naman Jain, Tianjun Zhang, Wei-Lin Chiang, Joseph E. Gonzalez, Koushik Sen and Ion Stoica
Under review at [ICLR 2024](#), Anaheim, California [preprint] [openreview]
Workshop Proceedings of [Neurips 2023](#), New Orleans, Louisiana [preprint]
6. Revisiting Prompt Engineering via Declarative Crowdsourcing
Aditya G Parameswaran, Shreya Shankar, Parth Asawa, **Naman Jain** and Yujie Wang
Proceedings of [ICDE 2023](#), Anaheim, California [preprint]
5. StaticFixer: From Static Analysis to Static Repair
Naman Jain, Shubham Gandhi, Atharv Sonwane, Aditya Kanade, Nagarajan Natarajan, Suresh Parthasarathy, Sriram Rajamani and Rahul Sharma
Preprint Available [preprint]
4. Jigsaw: Large Language Models meet Program Synthesis
Naman Jain, Skanda Vaidyanath, Arun Iyer, Nagarajan Natarajan, Suresh Parthasarathy, Sriram Rajamani and Rahul Sharma
Proceedings of [ICSE 2022](#), Pittsburgh, Pennsylvania [paper]
3. Learning Accurate Decision Trees with Bandit Feedback via Quantized Gradient Descent
Ajaykrishna Karthikeyan*, **Naman Jain***, Nagarajan Natarajan, and Prateek Jain
Proceedings of [TMLR 2022](#) [paper]
2. What's in a Name? Are BERT Named Entity Representations just as Good for any other Name?
Sriram Balasubramanian*, **Naman Jain***, Gaurav Jindal, Abhijeet Awasthi and Sunita Sarawagi
Workshop Proceedings of [ACL 2020](#), Virtual Conference [paper] [supplement]
1. On the Robustness of Human Pose Estimation
Naman Jain*, Sahil Shah*, Abhishek Sharma and Arjun Jain
Workshop Proceedings of [CVPR 2019](#), Long Beach, California [paper] [supplement]

* joint first authors

RESEARCH PROJECTS

RepoEval - Code Generation in the Wild

Nov'23 - Present

- Constructing an *executable* in-the-wild repository-context code generation benchmark with RAG setup
- Using program analysis (call-graphs and back-ward slicing) for providing reference-context in prompts

LiveCodeBench - Contamination Free and Holistic Code Evaluation

Oct'23 - Dec'23

- Performing periodically updated “live” benchmarking for code LLMs to evade data contamination
- Holistically evaluating LLMs for code generation, repair, execution, optimization, and test generation

CodeRefuse - Model Refusal for Coding Assistants

Jun'23 - Aug'23

- Constructed a high-quality model refusal dataset for user-queries using retrieval and in-context learning
- Transformed CWE docs from OWASP and MITRE into unsafe requests for data annotation pipeline

CodeQuality - LLM Assisted Code Cleaning

May'23 - Sep'23

- Proposed LLM-based refactoring for decomposing complex algorithmic programs into smaller modules
- Fine-tuning on “cleaned code” further improved code generation performance with explainable code

Parametric Template Distillation for API Summarization

Jan'22 - Apr'22

- Learning Parametric Templates (natural language with programmatic holes) for API summarization

Static Repair of Information Flow Vulnerabilities

Sep'21 - Nov'22

- Proposed a novel static-analysis-based approach for mining high-quality program-repair edits
- Repaired cross-site-scripting and unvalidated-call vulnerabilities in over 1000 repositories (90% recall)

Jigsaw – Combining Language Models with Program Synthesis

Dec'20 - Sep'21

- Proposed architecture for augmenting black-box models with program analysis and synthesis-based post-processing block that provides correctness guarantees and also allows learning from user feedback

DGT – Versatile Decision Tree Learning

Sep'20 - Oct'21

- Developed alternative conceptualization of decision tree problem that allows end-to-end gradient based learning while still performing competitive to SOTA tree methods and outperforming in bandit setting

Robustness in Natural Language Processing

Aug'19 - Jun'20

- Studied and designed training algorithms robust to synonym and named entity replacement attacks on tasks including sentiment analysis, grammar correction, coreference resolution, and question answering

Adversarial Examples in Human Pose Estimation

Aug'18 - Dec'18



- Performed image agnostic & dependent adversarial attacks evaluating crucial design choices such as direct regression vs heatmap, imagenet pretraining, using compositional human body structure, etc.

Interacting Humans Video Prediction

June 2019 - Dec 2020

- Introduced a novel conditional normalization approach for incorporating multi-person video context

SELECTED SOFTWARE AND OPEN SOURCE

- Human Pose Estimation - PyTorch (stats - ★>350, ♪>70) 
- Implemented `torch.randint` in PyTorch - open source contribution 

SCHOLASTIC ACHIEVEMENTS

- Awarded Undergraduate Research Award (**URA**) for Autumn 2018 (2 out of 121 students)
- Secured **All India Rank 36** in JEE Advanced 2016 among **0.15 million** candidates
- Awarded Gold Medal for being among top **35** students in Indian National Physics Olympiad (**INPhO**)
- Among the **top 300** in Indian National Chemistry and Astronomy Olympiads (**INChO & INAO**)
- Awarded KVPY Fellowship and NTSE Scholarship by Govt. of India

RELEVANT COURSES

- **Computer Science** - Automata Theory, Compilers, Parallelizing Compilers, Advanced Machine Learning, Learning Agents (RL), Computer Vision, Data Structures, Algorithms, Operating System
- **Mathematics** - Calculus, Linear Algebra, Differential Equations, Numerical Analysis, Discrete Maths

TECHNICAL STRENGTHS

Strong	Python, C, C++, C#, Racket, Prolog, Bash
Web	Flask, ASP.NET, Guicorn, Nginx, IIS, WebSockets, React, JSP
Tools	Git, L ^A T _E X, OpenGL, OpenMP, Gnuplot, Doxygen

TEACHING & MENTORSHIP

- Teaching Assistant for the course Automatic Speech Recognition under Prof. Preethi Jyothi
- Teaching Assistant for the course Quantum Mechanics under Prof. Aftab Alam
- Mentored 6 students for project on statistical face recognition organized by WnCC, IIT Bombay

REFERENCES

Prof. Sunita Sarawagi
Indian Institute of Technology, Bombay
[webpage](#) \diamond [email](#)

Dr. Sriram Rajamani
Microsoft Research, India
[webpage](#) \diamond [email](#)

Prof. Koushik Sen
University of California, Berkeley
[webpage](#) \diamond [email](#)

Dr. Nagarajan Natarajan
Microsoft Research, India
[webpage](#) \diamond [email](#)

Dr. Rahul Sharma
Microsoft Research, India
[webpage](#) \diamond [email](#)

Prof. Aditya Parameswaran
University of California, Berkeley
[webpage](#) \diamond [email](#)