Swapnil Gandhi

gandhis@stanford.edu • https://swapnilgandhi.com/

RESEARCH INTERESTS

I am interested in the system-side problems associated with training, deploying, and operationalizing deep learning models at scale.

EDUCATION

Ph.D. Candidate in Computer Science, *Stanford University*, GPA: 4.09/4.0 Jul 2022 – Present
- Advisor: Prof. Christos Kozyrakis
- Current Research Focus: Efficient fault recovery support for large scale model training

M.Tech. (Research), *Indian Institute of Science (IISc)*, GPA: 9.2/10.0 Aug 2017 – Jan 2020
- Advisor: Prof. Yogesh Simmhan
- Best M.Tech. (Research) Thesis (Honorable Mention) for "Distributed Programming Abstraction for Scalable Processing of Temporal Graphs"

B.Tech. in Computer Engineering, *Bharati Vidyapeeth Pune* Jul 2010 – Jun 2014
- Department Honors and Gold Medalist
- Thesis Title: Mutation Testing Tool for C Programs

PUBLICATIONS

[Papers & Posters available here.]

PEER-REVIEWED CONFERENCES

[1] Swapnil Gandhi, Christos Kozyrakis, "MoEtion: Efficient and Reliable Checkpointing for Mixture-of-Experts Models at Scale", *Under Review*

[2] Swapnil Gandhi, Mark Zhao, Athinagoras Skiadopoulos, Christos Kozyrakis "ReCycle: Pipeline Adaptation for the Resilient Distributed Training of Large DNNs", *In proceedings of the 30th ACM SIGOPS Symposium on Operating Systems Principles (**SOSP 2024**), Nov 2024.*
Acceptance Rate: $43/248 = 17.33\%$

[3] Anand Iyer, Mingyu Guan, Yinwei Dai, Rui Pan, Swapnil Gandhi, Ravi Netravali "Improving DNN Inference Throughput Using Practical, Per-Input Compute Adaptation", *In proceedings of the 30th ACM SIGOPS Symposium on Operating Systems Principles (**SOSP 2024**), Nov 2024.*
Acceptance Rate: $43/248 = 17.33\%$

[4] Swapnil Gandhi, Anand Padmanabha Iyer, "P^3: Distributed Deep Graph Learning at Scale", *In proceedings of the 15th USENIX Symposium on Operating Systems Design and Implementation (**OSDI 2021**), Jul 2021.*
Acceptance Rate: $31/165 = 18.78\%$

[5] Swapnil Gandhi, Yogesh Simmhan, "An Interval-centric Model for Distributed Computing over Temporal Graphs", *In proceedings of the 36th IEEE International Conference on Data Engineering (**ICDE 2020**), Dallas, Texas, April 2020.*
Acceptance Rate: $129/568 = 22.71\%$

PEER-REVIEWED POSTERS

[1] Swapnil Gandhi, "Wave: A Substrate for Distributed Incremental Graph Processing on Commodity Clusters", *2nd ACM Student Research Competition (SRC) at 27th Symposium on Operating Systems Principles (**SRC- SOSP 2019**), Ontario, Canada, Oct 2019.*
Received Bronze Medal, Student Research Competition (Graduate Category)

[2] Swapnil Gandhi, Sayandip Sarkar, Abhilash Sharma, Yogesh Simmhan, "Distributed Querying over Compressed Property Graphs", *Student Research Symposium at 24th IEEE International Conference on High Performance Computing, Data and Analytics (**HiPC 2017**), Jaipur, India, Dec 2017.*
Received Best Student Research Symposium Poster

AWARDS & HONORS

JUMP 2.0 ACE Best Paper — Nov 2024
For "ReCycle: Pipeline Adaptation for the Resilient Distributed Training of Large DNNs" in Quarter 3 2024.

Stanford Computer Science Student Service Award 2024 — Jun 2024

Stanford Computer Science Student Service Award 2023 — Jun 2023

NetApp Gold Medal for Best M.Tech (Research) Thesis (Honorable Mention), IISc — Jan 2022
For "Distributed Programming Abstraction for Scalable Processing of Temporal Graphs".

Selected to participate in The Cornell, Maryland, Max Planck Pre-doctoral Research School (CMMRS) 2020, Saarbrücken, Germany — Aug 2020

Bronze Medal, 2nd ACM Student Research Competition (Graduate Category), at SOSP — Oct 2019
For "Wave: A Substrate for Distributed Incremental Graph Processing on Commodity Clusters".

Won 12th IEEE International TCSC Scalable Computing (SCALE) Challenge — May 2019
For "Dynamic Scaling of Video Analytics for Wide-area Tracking in Urban Spaces".

Best Poster Award, 10th EECS Research Students Symposium, IISc Bangalore — Apr 2019
For "Distributed Processing Model For Temporal Graphs".

Invited to attend 3rd RIKEN R-CCS HPC Youth Workshop, Kobe, Japan — Feb 2019

Best Student Research Symposium Poster, IEEE HiPC, Jaipur, India — Dec 2017
For "Distributed Querying over Compressed Property Graphs".

Department Honors, Bharati Vidyapeeth, Pune — Jun 2014
For outstanding academic performance (Batch 2010 – 2014).

TCS Popular Student Project, Bharati Vidyapeeth, Pune — May 2014
For "Mutation Testing Tool for C Programs", Bachelors dissertation.

Best Undergraduate Project Award, TRDDC Annual Students Day, Pune — Apr 2014
For "Mutation Testing Tool for C Programs", Bachelors dissertation.

INDUSTRY EXPERIENCE

NVIDIA, Research Intern — Jan 2025 – Present

NVIDIA, AI Systems Engineering Intern — Jun 2024 – Sep 2024
Internship Mentor: Niket Agrawal
Researched techniques for joint placement and scaling for SLO-aware LLM serving in cloud

Microsoft Research India, Research Fellow — Jul 2021 – Sep 2022
Internship Mentor: Anand Iyer
Researched techniques for improving system-wide inference goodput for early-exit deep neural network using heterogeneous resources; led to a paper in SOSP 2024

Microsoft Azure R&D India, Software Engineer II — Mar 2021 – Jun 2021
Worked on query optimization and distributed execution strategies in SQL Server

Microsoft Research India, Research Intern — Sep 2020 – Mar 2021
Internship Mentor: Anand Iyer
Researched how model and data parallelism can be combined with independent graph partitioning for training Graph Neural Networks (GNNs) at scale; led to paper in OSDI 2021

Microsoft Research India, Research Intern — Mar 2020 – Aug 2020
Internship Mentors: Karthik Ramachandra, Bhargav Gulavani
Investigated and implemented query optimizer modifications to overcome performance regressions in scalar UDF inlined queries; shipped in SQL Server.

PubMatic India, Data Ops Engineer — Jun 2014 – Jul 2016
Worked on reporting and ad-hoc data processing pipelines using combination of Hadoop, Hive, and Pig.

TATA Research Development and Design Centre India, Research Intern — Sep 2013 – Apr 2014
Internship Mentors: Prasad Bokil, Ulka Shrotri, R. Venkatesh
Created prototype mutation testing tool for C programs; used by internal QA teams.

SERVICE

Co-Chair, Stanford CS PhD Admit Weekend	2023 – Present
Co-Organizer, Stanford CS Application Assistance Program (SASP)	2023 – Present
Artifact Evaluation Committee (AEC) Member, ACM EuroSys 2023	Aug 2022 – Oct 2022
Shadow PC Committee Member, ACM EuroSys 2022	Oct 2021 – Dec 2021
Shadow PC Extended Review Committee Member, ACM EuroSys 2021	Oct 2020 – Dec 2020
Artifact Evaluation Committee (AEC) Member, USENIX OSDI 2020	Aug 2020
Artifact Evaluation Committee (AEC) Member, ACM ASPLOS 2020	Dec 2019
Artifact Evaluation Committee (AEC) Member, ACM SOSP 2019	Aug 2019
Treasurer and General Secretary for IISc ACM Student Chapter	Apr 2019 – Mar 2020

TEACHING ASSISTANTSHIPS

CS 349D: Cloud Computing Technology, Stanford — Apr 2024

Graduate Teaching Assistant for CS 349D. Handled class discussion, homework assignments and helped with class projects (≈ 30 students).

DS 256: Scalable Systems for Data Science, IISc — Jan 2019

Graduate Teaching Assistant for DS 256. Handled weekly discussion sections, homework assignments and helped with class projects (≈ 25 students).

E0 261: Database Management Systems, IISc — Oct 2018

Covered papers on Google's Spanner and Apache Giraph. (≈ 40 students).

[*CV compiled on 2025-01-06*]