Shubham Gandhi

Website | LinkedIn | Google Scholar | GitHub | Email: srgandhi@andrew.cmu.edu | Phone: +1 412-954-8921

Education

Carnegie Mellon University - Language Technologies Institute (SCS) | GPA: 4.03/4.0
Ph.D. in Language Technologies — *Aug. 2026 - Present*
M.S. in Intelligent Information Systems — *Aug. 2024 - Dec. 2025*

Birla Institute of Technology and Science, Pilani — Aug. 19 - Jul. 23
B.E. in Computer Science, Minor in Data Science | *GPA* - 8.99/10 — *Goa, India*

Research Interests

LLM Agents | Efficiency & Failure Mitigation | Reliability & Security | Reasoning Alignment | AI for Code

Experience

Carnegie Mellon University | *Research Assistant* | *Prof. Carolyn Rosé, Prof. Michael Hilton* | Jan. 25 - Present

Secure and useful code generation for the Amazon Nova AI Challenge: utility-preserving reasoning defenses for code LLMs (Trusted AI '25, finalist) and adversarial red-teaming of code agents (Trusted Software Agents '26, team lead).

IBM Research | *Research Intern* | *Dr. Jason Tsay, Dr. Kiran Kate, Dr. Yara Rizk* | Summers '25, '26

Process supervision for long-horizon SWE agents: SWE-PRM for course-correcting agent trajectories at inference time ('25) and DRACO for fine-grained credit assignment with dynamic rubrics during agent training ('26).

TCS Research | *Research Fellow* | *Dr. Manasi Patwardhan, Dr. Gautam Shroff* | Jul. 23 - Jul. 24

LLM multi-agent systems for research automation: ResearchCodeAgent for codifying research methodologies, BudgetMLAgent for cost-efficient ML automation, and low-resource COBOL-to-Java translation.

Microsoft Research | *Research Intern* | *Dr. Nagarajan Natarajan, Dr. Sriram Rajamani* | Jun. 22 - Dec. 22

StaticFixer: repairing information-flow vulnerabilities in JavaScript by combining CodeQL static analysis with neural baselines (Codex, CodeT5); built the CodeInTheWild and PairedPrograms datasets.

Publications (*C = Conference, W = Workshop, TR = Tech. Report, * = Equal Contrib.*)

[W] **S. Gandhi***, S. Agarwal*, N. Kotalwar*, A. Naik - "Evaluating Code Slop in Long-Horizon Coding Agents", WAB @ COLM '26, Link

[C] **S. Gandhi***, Y. Xie*, A. Naik, R. Zhu, C. Rosé - "Steer, Don't Solve: Training Small Critic Models for Large Code Agents", EMNLP '26 (Oral), Link

[C] Y. Xie, E. Liu, G. Zhang, N. Kotalwar, **S. Gandhi**, S. Acharya, X. Wang, C. Rosé, G. Neubig, D. Fried - "Hybrid-Gym: Training Coding Agents to Generalize Across Tasks", ICML '26, Link

[W] **S. Gandhi**, J. Tsay, J. Ganhotra, K. Kate, Y. Rizk - "When Agents go Astray: Course-Correcting SWE Agents with PRMs", SEA @ NeurIPS '25, Link

[TR] A. Naik, A. Xie, A. Rao, A. Agarwal, **S. Gandhi**, M. Hilton, C. Rosé - "Secure and Useful Models are Reasonable: Aligning Code Models via Utility-Preserving Reasoning", Amazon Nova AI Proceedings '25, Link

[C] **S. Gandhi**, A. Naik, Y. Xie, C. Rosé - "An Empirical Study on Strong-Weak Model Collaboration for Repo-level Code Generation", EMNLP '25, Link

[W] **S. Gandhi**, D. Shah, M. Patwardhan, L. Vig, G. Shroff - "ResearchCodeAgent: An LLM Multi-Agent system for automated codification of research methodologies", AI4Research @ AAAI '25, Link

[C] **S. Gandhi**, M. Patwardhan, L. Vig, G. Shroff - "BudgetMLAgent: A Cost-Effective LLM Multi-Agent system for Automating Machine Learning Tasks", AI-ML Systems '24, Link

[W] **S. Gandhi**, M. Patwardhan, J. Khatri, L. Vig, R. K. Medicherla - "Translation of Low-Resource COBOL to Logically Correct and Readable Java leveraging High-Resource Java Refinement", LLM4Code @ ICSE '24, Link

Preprints (*P = Preprint, UR = Under Review, * = Equal Contrib.*)

[P] A. Oke, A. V. Duarte, R. Melo, **S. Gandhi**, N. Kotalwar, C. Khandor, D. Wang, C. Rosé, L. Li - "Ulysses: A Security Harness for Synthesizing Verifiable Web Exploits", arXiv preprint, '26, Link

[P] R. Melo, A. V. Duarte, A. Oke, **S. Gandhi**, N. Kotalwar, R. Abreu, L. Li, C. Rosé - "Self-Evolving Static Analysis via LLM Pseudo-Labeling", arXiv preprint, '26, Link

[P] N. Kotalwar, C. Borchers, **S. Gandhi**, P. S. Purkayastha, D. Yaron, C. Rosé - "Exploring Visual Access and Proactive AI Tutoring for Handwritten Mathematics", arXiv preprint, '26, Link

[UR] A. V. Duarte, A. Oke, R. Melo, **S. Gandhi**, N. Kotalwar, C. Khandor, D. Wang, A. L. Oliveira, C. Rosé, L. Li - "ABSENTIA: Detecting Broken Access Control Vulnerabilities in Web Applications", Under review, '26, Link

[UR] **S. Gandhi***, S. Goyal*, K. Kate, Y. Rizk - "DRACO: Fine-Grained Credit Assignment with Dynamic Rubrics for Long-Horizon Agent Training", Under review, '26, Link

[P] N. Jain, **S. Gandhi**, A. Sonwane, A. Kanade, N. Natarajan, S. Parthasarathy, S. Rajamani, R. Sharma - "StaticFixer: From Static Analysis to Static Repair", arXiv preprint, '23, Link

Select Research Projects

RepoQABench | *Instructors: Prof. Eric Nyberg, Prof. Teruko Mitamura* | *Report* | *Code* — Jan. 25 - May. 25

Created RepoQABench, a benchmark evaluating LLMs on multi-file repository-level code QA; led end-to-end data curation, QA generation, and taxonomy design.

LLMs for Code Quality Assessment | *Instructor: Prof. Graham Neubig* | *Report* | *Code* — Nov. 24 - Dec. 24

Developed a novel approach that leverages code modification metrics to classify code based on non-functional requirements like security and maintainability achieving 71.28% accuracy.

Teaching Experience

Teaching Assistant | *MCDS Capstone Planning Seminar - 11634* — Jan. 26 - Present

Teaching Assistant | *MCDS Capstone - 11632* — Aug. 25 - Dec. 25

Teaching Assistant | *MCDS Capstone Planning Seminar - 11634* — Jan. 25 - May 25

Teaching Assistant | *Deep Learning - CS F425* — Jan. 23 - May 23

Teaching Assistant | *Introduction to Data Mining - WILP* — Jan. 23 - May 23

Teaching Assistant | *Computer Programming - CS F111* — May 22 - Aug. 22

Workshop Instructor | *Introduction to R* — Nov. 21

Academic Service

Organizer | *LTI Student Research Symposium 2026*

Reviewer | *ARR May '26, EDM '26, SEA @ NeurIPS '25, IKDD CODS '25 Demo Track*

Technical Skills

Programming: Python, C/C++, Java, JavaScript, Bash
Frameworks & Infrastructure: PyTorch, Hugging Face Transformers, vLLM, veRL, LlamaFactory, CodeQL, Slurm, Docker, AWS

Relevant Coursework

Reinforcement Learning | Neural Code Generation | Advanced Natural Language Processing | Question Answering w/ Large Language Models | Machine Learning | Deep Learning | Meta Learning

Achievements

Amazon Nova AI Challenge – Trusted Software Agents Grant (Red Team) — 2026

Led CMU team for the Trusted Software Agents Track (red team); awarded a $250K research grant and monthly $150K AWS compute support for adversarial evaluation of code agents.

Amazon Nova AI Challenge Finalist – Trusted AI — 2025

Finalist (Top 4 defender teams globally); selected among 90+ teams and awarded a $250K research grant for developing secure code LLM systems.

Summer Internship Assistantship Award for Research — 2022

Among top 5 students out of 5,000+ to be awarded a grant for research on Brain-Inspired Neural Networks.

Merit Scholarship, Sem 1-2 — 2019-20

40% tuition fee waiver awarded to top 3%ile of students for academic excellence.