




> Sophia Kolak

☎ 845-729-4355 | @ sdk2147@columbia.edu |  LinkedIn |  GitHub |  Website | 📍 Manhattan, NY

> education

Carnegie Mellon University
MS, Software Engineering

Pittsburgh, PA
Fall 2021 - Fall 2024

Columbia University
BA, Computer Science

New York, New York
Fall 2017 - Spring 2021

> work & research experience

Applied Scientist & Researcher
Carnegie Mellon University

Pittsburgh, PA
2021 - Present

- Developed and validated a novel LLM-based bug detection and repair algorithm on 3,000+ real-world bugs, achieving a 49% accuracy improvement over existing methods.
- Designed and conducted scaling-law experiments for LLMs of source code (60M-12B parameters), elucidating the strong statistical correlation between a model's size and its ability to generate valid patches to buggy programs.

Quantum & Qiskit Developer (intern)
IBM Quantum

New York, NY
May 2022 - December 2022

- Researched and evaluated quantum algorithms for linear solvers, achieving up to a 30% improvement in graph calculation efficiency for complex workflows.
- Spearheaded the development of a key partnership with a major client, Ansys, leading collaborative quantum research initiatives that expanded and strengthened the IBM Quantum-Ansys relationship.
- Published research on the application of quantum algorithms to customer-relevant problems.

Software Engineering Researcher
Carnegie Mellon University

Pittsburgh, PA
Summer 2019, Summer 2020

- Conducted the first large-scale empirical study of the Robot Operating System (ROS) ecosystem, analyzing data from over 200,000 ROS packages and dependencies on GitHub and published findings at ICSME 2020.
- Built a SQL database to track and visualize the growth and collaboration patterns within the ROS ecosystem, processing over 2 million data points.
- Identified key collaboration hubs and contributors driving over 70% of ecosystem development, providing strategic recommendations for enhancing open-source collaboration among 740+ companies.

> publications

[1] Revisiting Unnaturalness for Automated Program Repair in the Era of Large Language Models
Yang, [Kolak](#), et al. Under Review [link](#)

[2] Evaluating Quantum Algorithms For Linear Solver Workflows
[Sophia Kolak](#), et al. International Supercomputing Conference 2023 [link](#)

[3] Patch Generation with Language Models: Feasibility and Scaling Behavior,
[Sophia Kolak](#), [Ruben Martins](#), [Claire Le Goues](#), [Vincent Hellendoorn](#), ICLR DL4C workshop 2022 [link](#)

[4] It Takes a Village to Build a Robot: An Empirical Study of the ROS Ecosystem,
[Sophia Kolak](#), [Afsoon Afzal](#), [Claire Le Goues](#), [Michael Hilton](#), [Chris Timperley](#), ICSME 2020 [link](#)

[5] Detecting Performance Patterns with Deep Learning,
[Sophia Kolak](#), SPLASH Companion 2020 [link](#)

[6] SHIRLEE: A Sharp-edge Handheld Identifier and Remover in Low-gravity Extravehicular
Environments, [Dada](#), [Ganeshan](#), [Groll](#), [Kolak](#), [Ravi](#), [Stein](#), [Wang](#), AIAA SciTech Forum 2021 [link](#)

> awards & achievements

AIAA Student Research Competition 2020, Best Paper
CRA Outstanding Undergraduate Researcher Award 2021, Finalist
SPLASH Student Research Competition 2020, 3rd Place
AWM Student Chapter 2020, Award for Scientific Excellence

> presentations & posters

Patch Generation with Large Language Models | [Poster](#)
ICLR DL4C Workshop, virtual 2022
It Takes a Village to Build a Robot | [Video](#) | [Slides](#)
ROScon, Macau 2019
Robotics Software Quality Panelist | [Video](#)
ROS World, virtual 2020
ROS Developers Podcast | [Video](#)
Featured Interview, 2020
Quantum Computing and Independence-Friendly Logic | [Program](#)
Deconstructing Hintikka, Dubrovnik 2020

> teaching, leadership & service

Association for Women in Mathematics | *President* | 2020-2021

- Organized weekly lectures and study groups within math department.

Art & Machine Learning | *Teaching Assistant* | 2023

- Assisted with lectures, demoed new generative AI tools, and held weekly office hours.

Computer Science Theory | *Teaching Assistant* | 2019-2020

- Ran office hours, graded assignments, managed piazza and held recitation.

Columbia Space Initiative | *Mission Co-Lead* | 2019-2020

- Led Columbia's team of undergraduate researchers in the NASA-SUITS design competition.

AWM Machine Learning Reading Group | *Lead* | 2020

- Ran weekly discussion on introductory topics in machine learning.

FIRST Robotics | *Volunteer* | 2022

- Volunteer judge for regional high-school robotics competition.

ROS World Program Committee | *Volunteer* | 2021

- Volunteer PC member for ROS World.

> skills

Programming: Python, Java, C/C++, MATLAB, R, SQL, GraphQL, Neo4j
Frameworks/Libraries: ROS, PyTorch, TensorFlow, Scikit-Learn, Qiskit

> relevant coursework

Program Analysis, Ethics & Robotics, Machine Learning, Programming Languages & Translators, Art & Machine Learning, Software Engineering for Startups, Advanced Software Engineering, Artificial Intelligence, Computation and the Brain, Senior thesis in Computer Science, Analysis of Algorithms, Modern Analysis, Quantum Computing, Computer Science Theory, Fundamentals of Computer Systems, Advanced Programming, Data Structures, Linear Algebra, Research in Computer Science I & II, Discrete Math, Multivariable Calculus