

# > Sophia Kolak

☎ 845-729-4355 | ✉ sdk2147@columbia.edu | 🔗 LinkedIn | 🐙 GitHub | 🌐 Website | 📍 Brooklyn, NY

## > education

---

Carnegie Mellon University  
MS, Software Engineering

Pittsburgh, PA  
Fall 2021 - Spring 2023

Columbia University  
BA, Computer Science

New York, New York  
Fall 2017 - Spring 2021

## > work & research experience

---

Software Engineering PhD Student  
Carnegie Mellon University

Pittsburgh, PA  
2021-23

- Developed an Automated Program Repair (APR) tool using large language model (LLM) entropy deltas, achieving a 49% improvement in Top-1 patch ranking accuracy compared to state-of-the-art machine learning tools.
- Investigated and established the correlation between LLM model size (ranging from 60M to 12B parameters) and test-pass accuracy, contributing to a deeper understanding of cross-language scaling laws, which informed the design of more efficient language models.

Quantum & Qiskit Developer (intern)  
IBM Quantum

New York, NY  
May 2022 - December 2022

- Developed a quantum encoding for mesh structures with up to 10,000 nodes, improving the efficiency via quantum random-access optimizers (QRAO), significantly enhancing computational performance.
- Documented and presented the commercial value of quantum research findings to Ansys (ISV), directly contributing to the expansion of their strategic partnership with IBM.

Software Engineering Researcher  
Carnegie Mellon University

Pittsburgh, PA  
Summer 2019, Summer 2020

- Led the first empirical study of the Robot Operating System (ROS) ecosystem, providing foundational insights into the growth and evolution of ROS across various industries.
- Mined and analyzed data from over 200,000 ROS packages on GitHub, tracking their dependencies and usage patterns over time, which enabled a comprehensive understanding of the ecosystem's development.
- Identified key collaboration hubs within the ROS community, pinpointing a core set of working groups and packages that drive the ecosystem, contributing to strategic recommendations for enhancing collaborative efforts among over 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

---

SPLASH Student Research Competition 2020, 3rd Place  
CRA Outstanding Undergraduate Researcher Award 2021, Finalist  
AIAA Student Research Competition 2020, Best Paper  
Micro-G NASA Challenge 2019, Winning Tool  
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](#)  
ROScn, 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, MySQL, GraphQL, Neo4j  
Frameworks/Libraries: ROS, PyTorch, TensorFlow, Scikit-Learn, Qiskit  
Languages: English (Native), Croatian (Elementary)

## > 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