SIMONS INSTITUTE
for the Theory of Computing

Simons Institute for the Theory of Computing
9th Annual Industry Day

Thursday, November 2, 2023

Table of Contents

Welcome

Industry Day Agenda

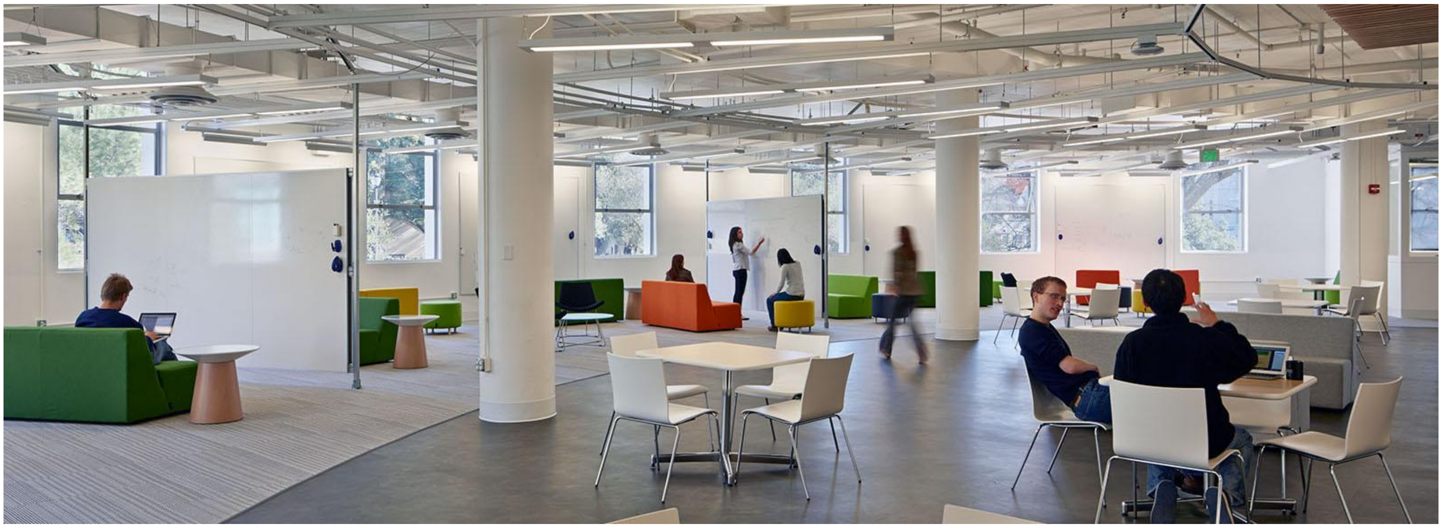
Preparation for Industry Day 2023

Program Presentations

Talks by Industry Partners

Research Fellows Lightning Talks

Special Program Sessions



Welcome

Dear friends,

Thank you for joining us for the 9th Annual Simons Institute Industry Day.

Today's event is once again a chance for our research fellows and industry scientists to take the stage, to share their work, and to uncover new connections and synergies in research.

Our industry partnerships are vital to the relevance of our research. They represent a variety of industries including computing software and hardware, computing analytics and security, and financial services and real estate. They provide a sounding board and reveal new perspectives and directions for research. As the demand for technology with richer functionality and enhanced robustness at larger scales continues to grow, the theory of computing and its principled approach is ever more relevant. We applaud our industry partners for supporting and being involved closely with core science. As the world becomes more complex and in some cases more fragile, together we can improve the human condition through joint discovery.

The support that we receive from our industry partners, and private individuals and foundations, ensures an active and productive second decade of research and discovery for the Institute.

To all, thank you for your visionary partnership, and your dedication to this important work.

Yours,

Venkat Guruswami, *Interim Acting Director and Senior Scientist*
Sandy Irani, *Associate Director*
Nikhil Srivastava, *Senior Scientist*

Industry Day Agenda

9:30 a.m. Coffee and Check-in

9:55 a.m. Introduction - Sandy Irani 5 mins

10:00 a.m. Program presentations

LLMs and Transformers: Aug. 14-18, 2023
Umesh Vazirani (15 minutes)
Logic and Algorithms In Database Theory and AI: Aug. 16- Dec. 15, 2023
Hung Ngo (3 minutes)
Data Structures and Optimization for Fast Algorithms: Aug. 16- Dec. 15, 2023
Aaron Sidford (3 minutes)
Error-Correcting Codes: Theory and Practice: Jan. 9 - May 10, 2024
Venkat Guruswami (3 minutes)
Quantum Algorithms, Complexity and Fault Tolerance: Jan. 9 - May 10, 2024
Sandy Irani (3 minutes)
Sublinear Algorithms: May 20 - Aug. 9, 2024
Jelani Nelson (3 minutes)
Modern Paradigms in Generalization: Aug. 26 - Dec. 13, 2024
Matus Telgarsky (15 minutes)
Cryptography 10 Years Later: Obfuscation, Proof Systems, and Secure Computation: May 19—Aug. 15, 2025
Abhishek Jain (15 minutes)

11:00 a.m. Break - 10 mins

11:15 a.m. Talks by Industry Partners

Unique Challenges and Opportunities in Working with Residential Real Estate Data
Takamitsu Tanaka, Head of Data and AI, Managing Director, Roc360

Lemur: Integrating Large Language Models in Automated Program Verification
Nina Narodytska, Senior Researcher, VMware

Introducing RelationalAI
Molham Aref, CEO, RelationalAI

Can Transformers Learn Algorithms? Mechanisms of Length Generalization
Preetum Nakkiran, Research Scientist, Apple

Market Algorithms for Autobidding
Gagan Aggarwal, Research Scientist, Research at Google

12:30 p.m. Lunch at the Institute

2:00 p.m Research Fellows - Lightning Talks

Join Algorithms Meet New Constraints: Scalability, Latency and Privacy
Xiao Hu

Vertex Connectivity in Poly-logarithmic Max-flows
Sorrachai Yingchareonthawornchai

Relational Programming
Remy Wang

Probabilistic Reasoning and Learning for Trustworthy AI
YooJung Choi

Non-monotonic convergence of gradient descent with large stepsize
Jingfeng Yu

Decoding the Maze: New Frontiers in Achieving Nash Equilibrium in AI Architectures
Manolis Vlatakis Gkaragkounis

How Close is Close Enough? Solving Linear Programs When Rounding Just Won't Cut It!
Bento Natura

How looking at a problem quantumly could solve it faster?
Hamoon Mousavi

From Robustness to Privacy and Back
Lydia Zakynthinou

3:00 p.m. Break

3:05 - 5:30 p. Special Sessions

Well-being, AI, and You: Developing AI-based Technology to Enhance our Well-being
Alon Halevy, Director, Facebook AI, Meta

Theory matters: how to build a successful company by trusting theoreticians
Molham Aref, CEO, Relational AI

DBSP: A unified (practical) theory of databases, streams, and incremental computation
Mihai Bidiu, Chief Scientist/Co-Founder and Leonid Ryzhyk, CTO/Co-Founder, Feldera

Challenges in Research on Practical Algorithms
Andrew Goldberg, Scientist

The Data Science behind Cruise's Robotaxi Fleet
Or Cohen, Senior Staff Data Scientist, Cruise

5:30 p.m. Reception at the Faculty Club

6:00 p.m. Dinner at the Faculty Club

Preparation for Industry Day 2023

Emcee

Sandy Irani, Simons Institute Associate Director, will run the show and manage the talks by program organizers and industry partners. Sandy will field questions for these sessions.

Moderator

Venkat Guruswami, Simons Institute Acting Interim Director and Senior Scientist, will manage the talks by research fellows. Venkat will field questions for this session.

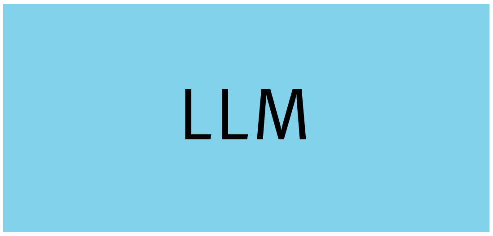
Recording Talks

Speakers who have signed release forms will be recorded, and their talks will be available after the event on the Simons Institute YouTube channel. If you would like to be recorded but have not yet received a release form, please contact Amy Ambrose: amyambrose@berkeley.edu

Point of Contact – Before & During the Event

Amy Ambrose
Sr. Development Director
amyambrose@berkeley.edu
+1 510 944 6674

Program Presentations



LLMs and Transformers

August 14 –18, 2023

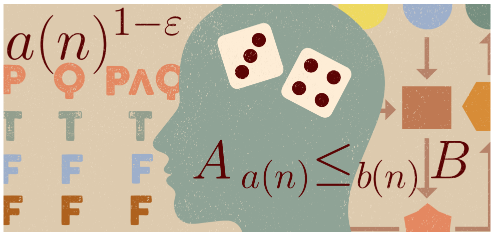
https://simons.berkeley.edu/workshops/large-language-models-transformers#simons-tabs

The goal of this workshop was to try to understand the ongoing revolution in transformers and large language models (LLMs) through a wide lens (including neuroscience, physics, cognitive science, and computation) in a relaxed setting that facilitates discussion, debate, and intellectual cross-pollination. The workshop touched on issues of fairness, trust, and alignment, and will seek to illuminate how industry and academia, and theory and systems, can collaborate.

The workshop talks are viewable here

Speakers: Scott Aaronson, Sanjeev Arora, Yasaman Bahri, Yonatan Belinkov, Yejin Choi, Paul Christiano, Miles Cranmer, Alyosha Efros, Surya Ganguli, Adam Kalai, Dan Klein, Yin Tat Lee, Jitendra Malik, Chris Manning, Steven Piantadosi, Colin Raffel, Sasha Rush, Pamela Samuelson, Ludwig Schmidt, Jacob Steinhardt, Ilya Sutskever, Josh Tenenbaum, Greg Valiant, and Diyi Yang.

Umesh V. Vazirani is a Research Director for Quantum Computing Simons Institute, and the Roger A. Strauch Professor of Electrical Engineering and Computer Science at the University of California, Berkeley, and the director of the Berkeley Quantum Computation Center. He received his BTech in Computer Science from MIT in 1981 and his PhD in Computer Science from Berkeley in 1985. Vazirani's research interests include computational learning theory, combinatorial algorithms, computational complexity theory and quantum computing. He received the 2012 Fulkerson Prize (with Sanjeev Arora and Satish Rao) for his work on approximation algorithms for sparsest cut.



Logic and Algorithms In Database Theory and AI

Aug. 16 – Dec. 15, 2023

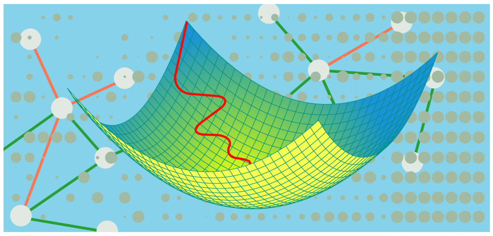
https://simons.berkeley.edu/programs/logic-algorithms-database-theory-ai

The last decade has brought a series of new theoretical results in database theory that lead to novel query evaluation and optimization algorithms, new information theoretic approaches to cardinality estimation, dichotomy results for incomplete and for probabilistic databases, tight bounds for the answer enumeration problem, new optimization techniques for tensor algebras, and extensions of logic semantics from the Boolean to arbitrary semirings. The common denominator of these results is their reliance on Logic, in that they apply to queries expressed in some logic. During the same time period, the complexity community on one hand, and the knowledge representation community on the other hand have produced results that inform and strengthen the developments in database theory. They include results in fine-grained complexity that establish sharp bounds for the model checking problem of first order formulas, with immediate application to the query evaluation problem in databases; results on polytope fooling that have potential to lead to improved algorithms for query evaluation on probabilistic databases; novel implementations of probabilistic logic programs based on circuits (e.g., problog); novel approaches to knowledge inference such as new statistical relational learning (e.g., PSL), and novel circuit-based knowledge compilation techniques (decision diagrams such as SDDs, Sum-Product Networks).

The program on Logic and Algorithms in Database Theory and AI brings together researchers from all three communities: database theory, complexity, and knowledge representation. The explicit goal of the program is to study the subtle interaction between logics and the algorithms that they inspire. The program will have four themes: Aspects of logical query evaluation; Fine-grained complexity through the lens of Logic; Logic-based aspects of circuit complexity and model counting; and Extensions of logics to semirings and aggregation.

This program is supported in part by Relational AI.

Hung Ngo was a professor at SUNY Buffalo from 2001 to 2015, when he left academia, and was with LogicBlox from 2015-2017. Hung joined RelationalAI on its founding day in 2017 and is currently Vice President of Research, and also leads the design and implementation of the query optimizer. Hung's current research interests are in query optimization and evaluation.He received best paper awards at COCOON 2008, ICDT 2019, PODS 2012, 2016, 2022.



Data Structures and Optimization for Fast Algorithms

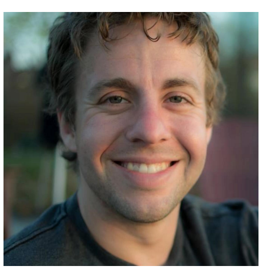
Aug. 16 – Dec. 15, 2023

https://simons.berkeley.edu/programs/data-structures-optimization-fast-algorithms

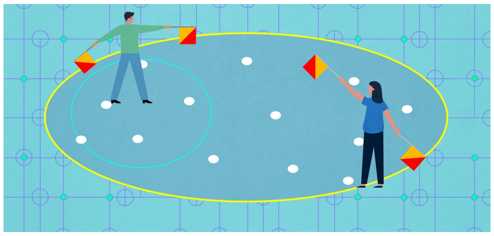
Over the past decade there has been a revolution in the theoretical foundations for obtaining provably fast algorithms for foundational problems in theoretical computer science. For problems ranging from maximum flow, to global minimum cut, to linear system solving, and beyond there have been breakthrough after breakthrough in improving asymptotic running times. Though classic theory had well-established the polynomial time solvability of these problems, the optimal running time for solving them had remained open; excitingly, recent work has shown that some of these problems can be solved in nearly linear time in broad regimes.

Beyond the foundational nature of these results and the modern motivations for faster algorithms, this suite of algorithmic advances is particularly exciting due to the wide range of tools they employ and the non-trivial interplay between them. These fast algorithms apply continuous optimization methods to solve combinatorial problems, dynamic data structures to solve static problems, sketching techniques to solve problems without memory limitations, and more. In some cases these faster algorithms have then in turn led to improvements for these tools, e.g. optimization techniques have been used in sketching routines and to yield improved dynamic graph algorithms. There have even been recent instances where a large suite of continuous, combinatorial, dynamic data structures, and sketching techniques are all combined for a faster algorithm for a single problem.

The central goal of the proposed program is to add further fuel to this budding area of research. The program will bring together experts in dynamic graphs, sketching, and optimization towards the common goal of furthering this confluence of advanced algorithmic techniques for breakthrough results. The bootcamp, workshops, and broader program activities will all be designed with the goal of advancing the foundations of each area, developing new connections between the areas, and making new advances at their intersection.



Aaron Sidford is an assistant professor in the departments of Management Science and Engineering and Computer Science at Stanford University. He received his PhD from the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology, where he was advised by Jonathan Kelner. His research interests lie broadly in the design and analysis of algorithms, optimization theory, and the theory of computation with an emphasis on work at the intersection of continuous optimization, graph theory, numerical linear algebra, and data structures. He is the recipient of a Microsoft Research Faculty Fellowship, a Sloan Research Fellowship, a NSF CAREER Award, an ACM Doctoral Dissertation Award honorable mention, and best paper awards in COLT, FOCS, and SODA for work in these areas.



Error-Correcting Codes: Theory and Practice

Jan. 9 – May 10, 2024

https://simons.berkeley.edu/programs/error-correcting-codes-theory-practice

Error-correcting codes are a fundamental tool for protecting data from noise. They appear naturally throughout computer science, electrical engineering, math, and physics. For that reason, they have a rich intellectual history in all these areas. While error-correcting codes have been studied since the 1950s, the last several years have seen a flurry of theoretical progress across all these areas. For example, in theoretical computer science, we have seen new constructions of list-decodable and locally decodable/testable codes; in electrical engineering and information theory, we have seen the emergence of polar codes, and of new results on the performance of classical codes on stochastic channels; in math, we have seen new combinatorial results and connections with algebraic geometry; and in physics, we have seen the recent development of good quantum LDPC codes.

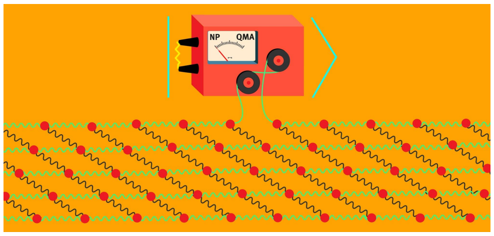
This flurry of theoretical progress has been matched with (and motivated by) a flurry of emerging applications. New technologies give rise to new theoretical problems, and solutions to those problems feed back into practice. Examples include applications in DNA storage, quantum computing, coding for distributed systems and distributed computing, and coding for emerging memory technologies. Error-correcting codes have also been finding new theoretical applications — for example, in complexity theory, pseudorandomness, and cryptography.

This program brings together researchers from different communities of theoretical research on error-correcting codes, including TCS, EE/IT, math, and physics, as well as researchers working on potential application areas. The aim is to further catalyze progress in this area and to lay out new directions for the field.

Venkatesan Guruswami is a senior scientist and currently Acting Interim Director at the Simons Institute for the Theory of Computing, and Professor of EECS and Mathematics at UC Berkeley. Venkat received his Bachelor's degree from the Indian Institute of Technology, Madras, in 1997, and his Ph.D. from MIT in 2001. He was a Miller Research Fellow at UC Berkeley and held faculty positions in Computer Science at the University of Washington and Carnegie Mellon University prior to moving to his current position in January 2022.

Venkat's research interests span many areas of theoretical computer science and related mathematics, including coding theory, approximate optimization, randomness in computing, and computational complexity.

Venkat has served the theory community in several leadership roles. He is the current Editor-in-Chief of the Journal of the ACM, and was previously Editor-in-Chief of the ACM Transactions on Computation Theory. He was program committee chair for CCC (2012), FOCS (2015), and ISIT (2018, co-chair), and recently president of the Computational Complexity Foundation. Venkat is a recipient of the 2023 Guggenheim Fellowship, a Simons Investigator award, the Presburger Award, Packard and Sloan Fellowships, the ACM Doctoral Dissertation Award, and an IEEE Information Theory Society Paper Award. He is a fellow of the ACM, AMS, and IEEE.



Quantum Algorithms, Complexity and Fault Tolerance

Jan 9–May 10, 2024

https://simons.berkeley.edu/programs/quantum-algorithms-complexity-fault-tolerance

Scaling up quantum computers to thousands of high-fidelity qubits and implementing fault-tolerant qubits is the next major challenge for experimental quantum computing. Theoretical challenges — such as the efficiency of protocols for fault-tolerant quantum computation, scalable proofs of quantumness, demonstrations of quantum advantage, and the development of quantum algorithms — will be critical to these practical efforts. These themes will all figure prominently in our program.

Many of this program's topics lie at the intersection of quantum computation and the theory of error-correcting codes. For example, recent constructions of quantum low-density parity-check (qLDPC) codes with optimal parameters show exciting potential for a new theory of quantum fault tolerance. However, many questions remain open, such as the efficiency of decoding algorithms for qLDPC codes, whether qLDPC codes can be locally testable, and how a fault-tolerant protocol can be implemented in a practical setting.

The program will also focus on quantum complexity theory and specifically on quantum Hamiltonian complexity. A central question here is the quantum PCP conjecture, which asks whether properties of many-body systems are hard to approximate. Recent progress on the quantum PCP conjecture has been a direct consequence of breakthroughs in qLDPC codes. Another central question is the area law conjecture, which states that ground (and low energy) states of physically relevant quantum systems have low entanglement and can be represented classically.

Quantum error correction, quantum complexity theory, and quantum cryptography are also playing an unexpected role in fundamental physics — namely, how to reconcile general relativity with quantum mechanics. The theories of entanglement and quantum error correction have led to progress on these questions, and there are strong indications that quantum computational complexity has a role to play in understanding the breakdown of effective field theory and reconciling the viewpoints of different observers in quantum gravity.

Another focus of this program is to further advance the ongoing revolution at the intersection of quantum computing and the theory of cryptography. One major theme is a reexamination of the foundations of cryptographic protocols in the presence of quantum adversaries. Another is to use a cryptographic leash to allow a classical verifier to carry out protocols with an untrusted quantum computer for tasks such as proofs of quantumness, certifiable randomness, verification of quantum memory, fully homomorphic quantum computation, and verification of quantum computation.

Sandy Irani is the associate director of the Simons Institute for the Theory of Computing. She received her PhD from UC Berkeley in 1991, after which she was a University of California President's Postdoctoral Fellow at UC San Diego. In 1992, she joined the faculty of UC Irvine, where she is currently a full professor. Much of her research has focused on algorithm design and analysis, with an emphasis on applications to computing systems. More recently, she has been working in quantum computation and quantum information science. Sandy's record of distinguished service includes two stints as chair of CS at UC Irvine, as well as an appointment as interim associate dean for undergraduate education; she also served as chair of the Ad Hoc Committee to Combat Sexual Harassment in the Theory of Computing Community.

SA

Sublinear Algorithms

May 20–Aug. 9, 2024

https://simons.berkeley.edu/programs/sublinear-algorithms

The ubiquity of large data sets has had a significant impact on the design of algorithms and has led to the emergence of computational models that capture various aspects of massive data computations, most notably that the data is too large to store or view in its entirety. Such models include streaming algorithms, sublinear time algorithms, sublinear query time algorithms with preprocessing, property testing algorithms, distributed and massively parallel algorithms. Though each of these models have somewhat different properties, what is common to them is that algorithms compute and make decisions based only on small local portions of the input.

In the *streaming setting,* big data streams go by and the algorithm has limited space in which to store information and intermediate computations; *sketching algorithms* compress the input in such a way that certain operations can be supported on the sketches; *sublinear time algorithms* approximate a parameter of the data after viewing a miniscule fraction of the data; *local algorithms* compute and make decisions on parts of the output after considering only a portion of the input; *property testing algorithms* quickly determine what properties are held by the data; *local distributed algorithms* are such that processors must output a decision based only a small number of rounds of interaction with only nearby processors, and hence based on inputs within a small local radius; the *Massively Parallel Computation* model of computation theoretically abstracts parallel cluster computing in practical settings where several commodity-grade computers interact to solve large-scale problems; *distribution testing algorithms* understand statistical properties of the data with sublinear samples in the size of the domain.

Sublinear algorithms in the above models have been studied for at least two decades, and a variety of very powerful tools have been developed in each. The past few years have seen a flurry of exciting developments and results, as well as the establishing of new connections between these sublinear models and common algorithmic and analysis techniques. Further, new paradigms have emerged, to incorporate new resources, constraints, or objectives to the sublinear setting: ranging from robustness, privacy and data erasure to interactive proofs and oracle-aided algorithms. The program will bring together researchers from the different communities of sublinear algorithms. After a bootcamp that will foster common language, there will be three workshops that focus on trends that cut across the communities: (1) *Sublinear graph simplification*, (2) *Sublinear algorithms with help*, and (3) *Extroverted sublinear algorithms*.

Jelani Nelson is Scientific Advisor at the Simons Institute for the Theory of Computing, and Professor of EECS at UC Berkeley. Nelson is interested in big data and the development of efficient algorithms. He joined the computer science faculty at Harvard University in 2013 and remained there until 2019 before joining UC Berkeley. He is known for his contributions to streaming algorithms and dimensionality reduction, including proving that the Johnson–Lindenstrauss lemma is optimal with Kasper Green Larsen, developing the Sparse Johnson-Lindenstrauss Transform with Daniel Kane, and an asymptotically optimal algorithm for the count-distinct problem with Daniel Kane and David P. Woodruff. He holds two patents related to applications of streaming algorithms to network traffic monitoring applications. Nelson was the recipient of an Office of Naval Research Young Investigator Award in 2015 and a Director of Research Early Career Award in 2016. He was awarded an Alfred P. Sloan Foundation Fellowship in 2017.

Nelson founded AddisCoder, a summer program teaching computer science and algorithms to high schoolers in Ethiopia, in 2011 while finishing his PhD at Massachusetts Institute of Technology. The program has trained over 500 alumni, some of whom have gone on to study at Harvard, MIT, Columbia, Stanford, Cornell, Princeton, KAIST, and Seoul National University, and to pursue PhDs in science and mathematics. Starting in 2022, Nelson also co-organized JamCoders, a summer algorithms and coding camp in Jamaica modeled on AddisCoder. Nelson also co-founded the David Harold Blackwell Summer Research Institute, which aims to increase the number of African-American students receiving PhDs in mathematics.

MPG

Modern Paradigms in Generalization

Aug. 26–Dec. 13, 2024

https://simons.berkeley.edu/programs/modern-paradigms-generalization

Generalization, broadly construed, is the ability of machine learning methods to perform well in scenarios outside their training data.

Despite being a well-developed field with a rich history, contemporary phenomena — in particular those arising from deep learning, most specifically large image and language models — are well beyond our current mathematical toolkit and vocabulary. It is not merely that analyses are too loose to be effective; rather, the settings have drastically evolved from the standard statistical setting of similar training and testing data, as the following examples illuminate: self-driving cars may need to navigate unfamiliar and even private or inaccessible roads; image generation software is expected to provide compelling images from essentially arbitrary input strings, with human operators indeed enjoying breaking the training data mold; AlphaFold and related software make protein predictions for species unrelated to those in their training set; the list goes on, without even scratching the surface of large language models and algorithmic tasks.

The goal of this program is to bring together remote and local researchers, both in academia and industry, as well as across mathematical and applied disciplines, with common goals of (a) organizing and crystallizing gaps between the theory and practice of generalization, and (b) sparking collaboration towards a concerted effort to close these gaps.

Matus Telgarsky is an assistant professor at the Courant Institute of Mathematical Sciences at New York University, specializing in deep learning theory. He was fortunate to receive a PhD at UCSD under Sanjoy Dasgupta. Other highlights include: co-founding, in 2017, the Midwest ML Symposium (MMLS) with Po-Ling Loh (while on faculty at the University of Illinois, Urbana-Champaign); receiving a 2018 NSF CAREER award; and organizing two Simons Institute programs, one on deep learning theory (summer 2019), and one on generalization (fall 2024).

Cryptography 10 years later: Obfuscation, Proof Systems, and Secure Computation

May 19—Aug. 15, 2025

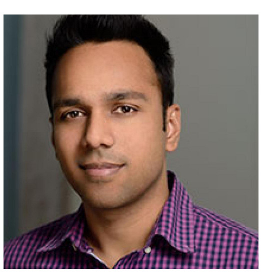
In an era where we have witnessed how data utilization can enable previously unthinkable progress across all domains of science (life, exact and social), healthcare, finance, and social connectivity, maintaining the privacy of sensitive personal data while leveraging it is at a critical juncture. At the same time, in the context of an upcoming international AI arms race, it has become clear that powerful machine learning algorithms will have to adhere to verification and auditing methods that ensure their correctness and trustworthiness even when executed in remote and adversarial environments.

Cryptographic tools have been crucial to both the goals of privacy and verification in the presence of adversaries. They enable safeguarding information in transit and in situ as well as when used in collaborative computation, amongst multiple distrustful parties who wish to extract utility from the union of their information, without sharing it fully. They also enable efficient protocols that verify the correctness of remote program executions and use of data, without replication or full knowledge of the internal details of the remote programs being verified.

This program will focus on obfuscation, secure systems, and secure computation, and extend existing theoretical cryptographic protocols and models to the context of machine learning, larger-scale multi-party collaborations on private data when adversarial parties are present, and new untrusted execution environments.

This program is poised to achieve breakthroughs on several key challenges within cryptography. Sample challenges include program obfuscation under well-established post-quantum assumptions; the existence of succinct and/or zero-knowledge proof systems with small computational overhead for general NP computations; fully homomorphic encryption based on new hard tasks for quantum computers to add on to the arsenal of existing lattice problems-based fully homomorphic encryption; and improving both on the asymptotic and the concrete efficiency of secure multi-party protocols for computations that naturally arise in ML settings.

The societal impact of solving these scientific challenges will be immense at this point in time, in recognition of the host of privacy and security risks that emerge with the proliferation of algorithms for data analysis and machine learning technologies. It is also imperative that in the future, algorithms of widespread usage will adhere to verification and auditing methods, to ensure their correctness and trustworthiness even when executed in remote and adversarial environments. This program will accelerate the development of cryptographic methods to ensure both the governance of algorithm execution and the privacy of data, for the benefit of society.



Abhishek Jain is a Scientist in NTT Research's CIS Lab, where his focus is on cryptography, computer security and privacy, and related topics in theoretical computer science. An Associate Professor of Computer Science at Johns Hopkins University, he completed his postdoctoral work at the Massachusetts Institute of Technology and his Ph.D. at University of California, Los Angeles, where he received the Outstanding Graduating Student Award. Additionally, he received his B.S. in Computer Science from UCLA, and a B.Tech in Computer Science and Engineering from the Indian Institute of Technology. To date, his research has received best-paper awards at EUROCRYPT 2021 and 2011, an NSF CAREER award, and JP Morgan Faculty Award.

Industry Partners

Unique Challenges and Opportunities in Working with Residential Real Estate Data

Takamitsu Tanaka ("Taka") is Head of AI Research and Managing Director at Roc360. He is a team leader specializing in data strategy, business intelligence, applied machine learning, and end-to-end custom project execution. Taka is at his best when collaboratively grappling with ambiguous, complex problems, and when sharing stories through data. Taka holds a B.S. in Biophysics (magna cum laude) from Cornell University, an M.S. in Physics from Penn, and a Ph.D. in Astronomy from Columbia University. His dissertation was entitled: "Mergers of Supermassive Black Hole Binaries in Gas-rich Environments: Models of Event Rates and Electromagnetic Signatures".

Lemur: Integrating Large Language Models in Automated Program Verification

The demonstrated code-understanding capability of LLMs raises the question of whether they can be used for automated program verification, a task that typically demands high-level abstract reasoning about program properties that is challenging for verification tools. We propose a general methodology to combine the power of LLMs and automated reasoners for automated program verification. We formally describe this methodology as a set of derivation rules and prove its soundness. We instantiate the calculus as a sound automated verification procedure, which led to practical improvements on a set of synthetic and competition benchmarks.

Nina Narodytska is a research scientist in the VMware research group. Prior to VRG, she was a researcher at Samsung Research America. She completed postdoctoral studies in the Carnegie Mellon University School of Computer Science and the University of Toronto. She received her PhD from the University of New South Wales in 2011. Nina works on developing efficient search algorithms for decision and optimization problems. She was named one of "AI's 10 to Watch" young researchers in the field of AI. She also received an Outstanding Paper Award at AAAI 2011 and an outstanding program committee member award at the Australasian Joint Conference on Artificial Intelligence 2012.

Introducing RelationalAI

Molham Aref is the Chief Executive Officer of RelationalAI. He has more than 28 years of experience in leading organizations that develop and implement high value machine learning and artificial intelligence solutions across various industries. Prior to RelationalAI he was CEO of LogicBlox and Predictix (now Infor), Optimi (now Ericsson), and co-founder of Brickstream (now FLIR). Molham held senior leadership positions at HNC Software (now FICO) and Retek (now Oracle).

Can Transformers Learn Algorithms?
Mechanisms of Length Generalization

Large language models exhibit surprising emergent generalization properties, yet also struggle on many simple reasoning tasks such as arithmetic and parity. To clarify the scope of Transformers' generalization abilities, we focus on the specific setting of length-generalization for algorithmic tasks. Here, we propose a unifying framework which leverages the recently introduced RASP language (Weiss et al 2021), which is a programming language tailor-made for the Transformer's computational model. Specifically, we introduce the RASP-Generalization Conjecture: Transformers tend to length-generalize on a task if there exists a short RASP program that solves the task for all input lengths. This simple conjecture remarkably captures most known instances of length generalization on algorithmic tasks. We also leverage our insights to give new scratchpad formats which allow Transformers to length-generalize easily on traditionally hard tasks (such as parity and addition). Finally, on the theoretical side, we give a simple example where the "min-degree-interpolator'" model of learning from Abbe et al. (2023) does not correctly predict Transformers' out-of-distribution behavior, but our conjecture does. Joint work with Hattie Zhou, Arwen Bradley, Etai Littwin, Noam Razin, Omid Saremi, Josh Susskind, Samy Bengio.

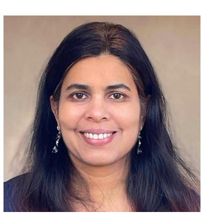
Preetum Nakkiran is a Research Scientist at Apple, in the Machine Learning Research Group. His research builds conceptual tools for understanding learning systems, including deep learning--- using both theory and experiment as appropriate. His past works include Deep Double Descent, the Deep Bootstrap Framework, and Distributional Generalization. Preetum obtained his PhD in Computer Science at Harvard, advised by Boaz Barak and Madhu Sudan. During his PhD, he co-founded the Harvard ML Foundations Group, and co-ran the corresponding seminar series. He did his postdoc at UCSD with Misha Belkin, and was part of the NSF/Simons Collaboration on the Theoretical Foundations of Deep Learning. He has also worked with Google Research and OpenAI, and is the prior recipient of the Google PhD Fellowship and the NSF GRFP. Preetum did his undergraduate work in EECS at UC Berkeley.

Research at Google

Market Algorithms for Autobidding

Over the past few years, more and more Internet advertisers have started using automated bidding for optimizing their advertising campaigns. Advertisers using automated bidding have an optimization goal (e.g. to maximize conversions), and some constraints (e.g. a budget or an upper bound on average cost per conversion), and the automated bidding system optimizes their auction bids on their behalf. In this talk, we will touch on fundamental questions that arise in this setting. How should an autobidder bid to optimize their goals? Do optimal bidding algorithms depend on the underlying auction? What happens when all advertisers adopt optimal autobidding? And what happens when bidders are optimizing across platforms?

The talk is based on joint work with Ashwinkumar Badanidiyuru, Aranyak Mehta, Andres Perlroth, Ariel Schvartzman, Junyao Zhao and Mingfei Zhao.



Gagan Aggarwal is a Research Scientist at Google, where she co-leads the Market Algorithms research team. Her research interests are in Algorithmic Game Theory and Approximation Algorithms, as well as their application to online marketplaces. She received her Ph.D. in Computer Science in 2005 from Stanford University, and her BTech in Computer Science in 2000 from IIT Delhi.

Research Fellows

Vertex Connectivity in Poly-logarithmic Max-flows

The vertex connectivity of an undirected graph is the smallest number of vertices whose removal disconnects the graph, which roughly measures the robustness of a graph against node failures. In this talk, I briefly present a celebrated "polylog max-flow" algorithm for computing vertex connectivity, which is the first improvement in the last 20 years. This gives an almost linear time vertex connectivity algorithm by using the recent breakthrough in the max-flow algorithm [CKLPPS'22]. Therefore, we settle the open question asked by Aho, Hopcroft, and Ullman [AHU'74] nearly 50 years ago (up to a sub-polynomial factor).

Joint work with Jason Li, Danupon Nanongkai, Debmalya Panigrahi, Thatchaphol Saranurak.

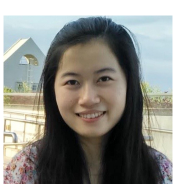
Sorrachai Yingchareonthawornchai
Research Fellow
Program: Data Structures and Optimization for Fast Algorithms
Dates of Visit: Aug. 16–Dec. 15, 2023

Sorrachai Yingchareonthawornchai is a doctoral student in theoretical computer science at Aalto University under the supervision of Parinya Chalermsook and Danupon Nanongkai. He received a MS in Computer Science from Michigan State University, and MEng in Computer Engineering from Chulalongkorn University. Previously, he received a BEng in Computer Engineering from Chulalongkorn University.

His research interests include fast graph algorithm design, combinatorial optimization, and extremal combinatorics.

Join Algorithms Meet New Constraints: Scalability, Latency and Privacy

Query evaluation is a fundamental problem in database systems, studying how to compute efficiently answers to questions specified in SQL. Join comes as the core operator, since many real-world analytical queries rely on the join operator to link information stored in different tables. The need to process and analyze big data has invigorated this long-running research area with new challenges. Firstly, data has grown tremendously both in size and complexity beyond a single machine's capability, which has resulted in massively parallel systems for data processing. How to scale join evaluation with thousands of machines in parallel? Secondly, data is generated at very high speeds. How to deliver timely answers to join queries with low latency? Thirdly, data is typically integrated from multiple sources and contains sensitive information, which has given rise to increasing privacy concerns. How to evaluate join queries without revealing sensitive information in the data? In this talk, I will discuss the challenges of establishing a strong theoretical foundation for join algorithms to address the new challenges brought by emerging data applications, and show some promising research directions in the interaction of scalable, low-latency and private join evaluation.



Xiao Hu
Research Fellow
Program: Logic and Algorithms in Database Theory and AI
Dates of Visit: Aug. 16–Dec. 15, 2023

Xiao Hu is an incoming Assistant Professor in the Cheriton School of Computer Science at University of Waterloo. She got her Ph.D. from HKUST, supervised by Ke Yi. Before that, she received her bachelor's degree from Tsinghua University. She has broad interest in database theory and algorithms.

Relational Programming

The practice of programming has undergone several paradigm shifts, from imperative, to object-oriented, then to functional. I believe the programming paradigm of the future is relational.

Remy Wang
Research Fellow
Program: Logic and Algorithms in Database Theory and AI
Dates of Visit: Aug. 16–Dec. 15, 2023

Remy Wang is a PhD student at University of Washington. He works at the intersection of Databases and Programming Languages. He develops languages, theory and tools to help data analysts conduct complex tasks efficiently. His research has been recognized with best paper awards from PODS, POPL and OOPSLA.

Probabilistic Reasoning and Learning for Trustworthy AI

There is an increasing interest and need for artificial intelligence and machine learning systems that are fair, robust, interpretable, and generally trustworthy. These properties about a model's behavior often depend on the underlying distribution of the world where it will be deployed, with various uncertainties that arise when deploying a system in the real world. In this talk, I will discuss how tractable probabilistic reasoning and learning can provide a unified language and framework under which we can develop trustworthy AI systems and present some recent work in algorithmic fairness in this direction.

YooJung Choi
Research Fellow
Program: Logic and Algorithms in Database Theory and AI
Dates of Visit: Aug. 16–Dec. 15, 2023

YooJung Choi is an assistant professor of computer science at Arizona State University. Her research interests lie in probabilistic machine learning, tractable probabilistic modeling and inference, knowledge representation and reasoning, as well as trustworthy artificial intelligence (algorithmic fairness, interpretability). Previously, she completed her PhD in computer science at UCLA.

Non-monotonic convergence of gradient descent with large stepsize

Gradient descent (GD) is a fundamental algorithm for optimizing machine learning models. However, certain observed behaviors of GD cannot be fully explained by classic optimization theory. For example, the training loss induced by GD often oscillates locally yet still converges in the long run. In the talk, I will provide insight into this phenomenon by considering logistic regression with separable data. I will show that GD can minimize the loss in the long run even when equipped with a large constant stepsize that violates the descent lemma.

Jingfeng Wu
Research Fellow
NSF/Simons Collaboration on the Theoretical Foundations of Deep Learning

Jingfeng Yu is a Postdoctoral Researcher at the Simons Institute for the Theory of Computing at UC Berkeley, hosted by Peter Bartlett and Bin Yu as a part of the NSF/Simons Collaboration on the Theoretical Foundations of Deep Learning. Jingfeng is interested in algorithms, machine learning, optimization, and statistical learning theory, especially in deep learning theory. Before coming to Berkeley, Jingfeng completed his PhD in Computer Science at Johns Hopkins University under the supervision of Vladimir Braverman. He received his MS and BS in Mathematics from Peking University.

How Close is Close Enough? Solving Linear Programs When Rounding Just Won't Cut It!

Imagine inching ever closer to a target, each attempt more precise than the last, yet still not quite hitting the mark. This encapsulates our dilemma with Linear Programs (LPs) when traditional rounding techniques are not up to par. In this lightning talk, we'll navigate the maze of challenging LPs and spotlight that, at least for LPs with just two variables per inequality, we have found a way to efficiently zero in on our goal.

Bento Natura
Research Fellow
Program: Data Structures and Optimization for Fast Algorithms
Dates of Visit: Aug. 16–Dec. 15, 2023

Bento Natura is a Postdoctoral Researcher at the Georgia Institute of Technology. He received his PhD in the Department of Mathematics at the London School of Economics under supervision of László Végh. Both his Bachelor's and Master's degree in Mathematics were received from the University of Bonn under supervision of Stephan Held and Jens Vygen. His research interests are in Operational Research, Combinatorial Optimization, Convex Optimization and Game Theory.

How looking at a problem quantumly could solve it faster?

In this talk I will tell you how we may solve hard computational problems if we change our perspective and look at the problem from the lens of quantum mechanics.

Hamoon Mousavi
Research Fellow
Program: Summer Cluster on Quantum Computing

Hamoon Mousavi is a PhD student at the Computer Science Department at Columbia University, advised by Henry Yuen. He is interested in applications of representation theory and semidefinite programming methods in entanglement and non-locality. He completed his master's in combinatorics and optimization, with a focus on quantum information at the University of Waterloo.

Decoding the Maze: New Frontiers in Achieving Nash Equilibrium in AI Architectures

In the contemporary realm of machine learning, applications such as adversarial attacks and multi-agent reinforcement learning can be viewed as intricate multi-agent games, where Nash equilibria represent optimal system states. Amidst the evident non-convex loss landscape of these games, a latent convex structure emerges, offering a potential route to convergence. In this talk, we will introduce a pioneering first-order method adept at leveraging this concealed structure, ensuring convergence to a Nash equilibrium in such non-convex games.

Merging traditional game theory with neural network dynamism, our work addresses the obfuscated nature of AI architectures by navigating the maze of control and latent variables, charting a path for convergence in hidden monotone games.

Emmanouil (“Manolis”) Vasileios Vlatakis Gkaragkounis

Research Fellow

NSF/Simons Collaboration on the Theoretical Foundations of Deep Learning

Manolis is a Simons-FODSI Post-doctoral fellow at UC Berkeley, working with Prof. Michael Jordan. He received his Ph.D. in Computer Science from Columbia University, under the supervision of Prof. Mihalis Yannakakis and Prof. Rocco Servedio.

His main research interests lie in the areas of Beyond Worst-Case Analysis of Algorithms, Theoretical Foundations of Machine Learning, Algorithmic Game Theory, Optimization, and Dynamical Systems. Before being a postdoc at UC Berkeley, he was a Research Fellow at the Simons Institute for Theory of Computing in the Learning and Games program, and at the "Athena" Research & Innovation Center in Athens, Greece. He received a Bachelor's and M.Sc. degree in Electrical and Computer Engineering from one of Greece's "Grand Schools", the ECE Department of the National Technical University of Athens, where he was advised by Prof. Dimitris Fotakis.

From Robustness to Privacy and Back

We study the relationship between two desiderata of algorithms in statistical inference and machine learning --- differential privacy and robustness to adversarial data corruptions. Their conceptual similarity was first identified by Dwork and Lei (STOC 2009), who observed that private algorithms satisfy robustness. Complementing this view, our work gives the first black-box transformation that converts any adversarially robust algorithm into one that satisfies differential privacy. We show that this implies that for any low-dimensional task, the optimal error rate for ε-differentially private estimators is essentially the same as the optimal error rate for estimators that are robust to adversarially corrupting $1/\varepsilon$ training samples. Our transformation gives us a systematic way to obtain new optimal private estimators by using their robust counterparts as a black box, which we demonstrate for several high-dimensional tasks including Gaussian (sparse) linear regression and PCA. Joint work with Hilal Asi and Jonathan Ullman (ICML 2023, https://arxiv.org/abs/2302.01855)

Lydia Zakynthinou

Postdoctoral Fellow

UC Berkeley EECS

Lydia Zakynthinou is a postdoctoral researcher at UC Berkeley. She received her PhD from Khoury College of Computer Science at Northeastern University, and her ECE diploma and Master's degree from NTUA and NKUA, respectively. She received a Facebook Fellowship in 2020 and the Khoury PhD Research Award in 2022. Lydia's research interests lie on the theoretical foundations of responsible and reliable machine learning. In particular, her research focuses on learning under privacy constraints and obtaining guarantees of validity and generalizability via algorithmic stability.

Special Sessions

Well-being, AI, and You: Developing AI-based Technology to Enhance our Well-being

Many applications claim to enhance our well-being, whether directly by aiding meditation and exercise, or indirectly, by guiding us to our destinations, assessing our sleep quality, or helping us manage our daily tasks. However, the truth is that the potential of technology to improve our well-being often eludes us, and this is happening at the dawn of an era where AI is supposed to usher in a new generation of personalized assistants. Presently, we find ourselves more distracted than ever, devoting excessive time to pondering life's minutiae, and struggling to fully embrace the present moment.

Part of the reason that our well-being is not benefiting fully from technology is the fact that each of these apps focuses on a specific aspect of well-being, lacking coordination with other apps. This situation is reminiscent of the early days of computer programming when each program interacted directly with the computer's hardware. Drawing from this analogy, this talk will begin by describing a set of mechanisms that can facilitate better cooperation between well-being applications, effectively proposing an operating system for well-being. This operating system comprises a data repository, referred to as a personal timeline, which captures your past experiences and future aspirations. It also includes mechanisms for utilizing your personal data to provide improved recommendations and life plans, and, lastly, a module to assist in nurturing and navigating crucial relationships in your life.

The second half of the talk will delve into the technical challenges involved in building the components of the operating system. In particular, we will focus on the creation of your life experiences timeline from the digital data you create on a daily basis. In this context, we will identify opportunities for language models to be a core component on which we build systems for querying personal timelines and for supporting other components of the operating system. In particular, the challenge of answering questions about your timeline raises important challenges in the intersection of large language models and structure data.

Alon Halevy is a director at Meta's Reality Labs Research, where he works on Personal Digital Data, the combination of neural and symbolic techniques for data management and on Human Value Alignment. Prior to Meta, Alon was the CEO of Megagon Labs (2015-2018) and led the Structured Data Group at Google Research (2005-2015), where the team developed WebTables and Google Fusion Tables. From 1998 to 2005 he was a professor at the University of Washington, where he founded the database group. Alon is a founder of two startups, Nimble Technology and Transformic (acquired by Google in 2005). Alon co-authored two books: *The Infinite Emotions of Coffee* and *Principles of Data Integration*. In 2021 he received the Edgar F. Codd SIGMOD Innovations Award. Alon is a Fellow of the ACM and a recipient of the PECASE award and Sloan Fellowship. Together with his co-authors, he received VLDB 10-year best paper awards for the 2008 paper on WebTables and for the 1996 paper on the Information Manifold data integration system.

Theory matters: how to build a successful company by trusting theoreticians

Molham Aref is the Chief Executive Officer of RelationalAI. He has more than 28 years of experience in leading organizations that develop and implement high value machine learning and artificial intelligence solutions across various industries. Prior to RelationalAI he was CEO of LogicBlox and Predictix (now Infor), Optimi (now Ericsson), and co-founder of Brickstream (now FLIR). Molham held senior leadership positions at HNC Software (now FICO) and Retek (now Oracle).

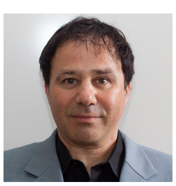
DBSP: A unified (practical) theory of databases, streams, and incremental computation

Mihai Budiu is chief scientist and co-founder at Feldera. He has a Ph.D. in CS from Carnegie Mellon University. He was previously employed at VMware Research, Barefoot Networks, and Microsoft Research. Mihai has worked on reconfigurable hardware, computer architecture, compilers, security, distributed systems, big data platforms, large-scale machine learning, programmable networks and P4, data visualization, and databases; four of his papers have received “test of time” awards. He has also received two technology transfer awards.

Leonid Ryzhyk is CTO and co-founder at Feldera. He holds a PhD. in computer science from the University of New South Wales. Before co-founding Feldera he worked as a researcher at NICTA, University of Toronto, Carnegie Mellon University, Samsung Research America, and most recently at VMware. Leonid has published dozens of research papers on operating systems, programming languages, formal verification, software-defined networks, and databases.

Challenges in Research on Practical Algorithms

In the modern world, computers and computational devices perform more and more tasks. We need to solve new problems and scale algorithms to bigger instances. This often requires new and improved algorithms. Both academia and industry faces challenges in developing such algorithms and putting them in production. In this talk we discuss some of these challenges and ways to mitigate them.

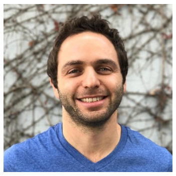


Andrew Vladislav Goldberg is an American computer scientist working primarily on design, analysis, and experimental evaluation of algorithms. He has also worked on mechanism design, computer systems, and complexity theory. Goldberg completed his undergraduate studies at the Massachusetts Institute of Technology, a master's degree at the University of California, Berkeley, and his PhD at MIT with a prestigious Hertz Fellowship. His thesis was on the Efficient graph algorithms for sequential and parallel computers and supervised by Charles E. Leiserson.

After completing his PhD, Goldberg was on the faculty of Stanford University and worked for NEC Research Institute, Intertrust STAR Laboratories, and Microsoft Research Silicon Valley Lab. He joined Amazon.com in 2014. Goldberg is best known for his research in the design and analysis of algorithms for graphs and networks, and particularly for his work on the maximum flow problem and shortest path problem, including the discovery of the push–relabel maximum flow algorithm. He also worked on algorithmic game theory, where he was one of the first scientists to study worst-case mechanism design.

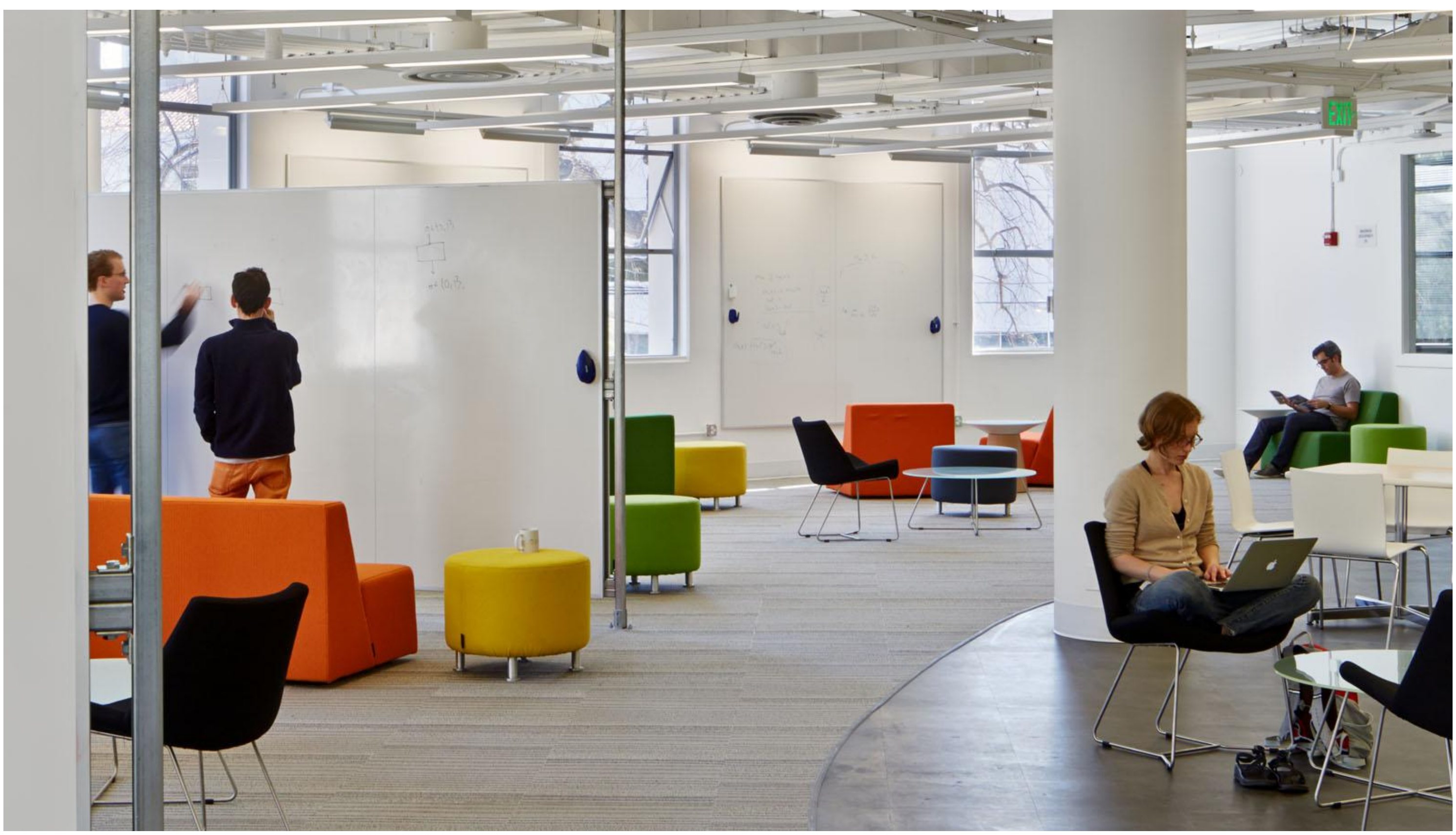
The Data Science behind Cruise's Robotaxi Fleet

In my talk I will give a quick overview of the unique challenges facing the Data Science team at Cruise. I will briefly talk about measuring the safety of our robotaxis on the road and in simulation. I will focus much of the talk on optimizing their operation of our fleet of robotaxis in areas such as vehicle routing, positioning, charging and ride matching. I will emphasize how operating a one-sided market where vehicles are fully controlled by the operators creates different challenges than what Uber and Lyft face.



Or Cohen, Senior Data Scientist, Cruise

Or Cohen is Senior Staff Data Scientist at Cruise. Prior to joining Cruise, Cohen was Staff Scientist at Lyft and before that, Head of Engineering and AI at modeai. Cohen holds a Ph.D. in Physics from Weizmann Institute of Science.



Thank you for your partnership!

Please join us for these coming programs:

Error-Correcting Codes: Theory and Practice ***January 9 – May 10, 2024***
Quantum Algorithms, Complexity and Fault Tolerance ***January 9 – May 10, 2024***
Sublinear Algorithms ***May 20 – August 9, 2024***
Theoretical Foundations of Computer Systems (extended reunion for original participants and industry partners)
July 1 – August 9, 2024
Modern Paradigms in Generalization ***August 26 – December 13, 2024***
AI, Psychology, and Neuroscience ***June 3 – June 28, 2024***
Special Year on Large Language Models and Transformers: Part I ***Aug. 26 – Dec. 13, 2024***
Special Year on Large Language Models and Transformers: Part 2 ***Jan. 14 – May 2, 2025***
New: Cryptography 10 Years Later: Obfuscation, Proof Systems, and Secure Computation
May 20 – August 9, 2025 (link TBD)

Join our Industry Partner community
The Simons Institute collaborates with industry to inform our research with real-world examples of theory in practice. Benefits of industry partnerships include participation in research program workshops and in industry-oriented events such as our annual Industry Day, research fellow namings, membership in our Industry Advisory Council, and many other opportunities to connect with our international community of distinguished scientists and students. Please contact Amy Ambrose (amyambrose@berkeley.edu) for more information on becoming a partner.

Simons Institute Matching Fund
Simons Institute is supported by grants and gifts from private foundations, industry affiliates, individuals, and governmental agencies. Currently, all gifts and pledges committed to the Simons Institute, including industry contributions, will be matched 1-to-1 with a gift to the Simons Institute's endowment, up to $25 million. This generosity will provide vital support to our community of senior researchers, postdoctoral fellows, and graduate students.

Please contact us for more information on supporting the Simons Institute, either as an industry affiliate or as an individual.

Simons Institute contacts:

Venkat Guruswami, Acting Interim Director, simonsdirector@berkeley.edu

Sandy Irani, Associate Director, simonsassociatedirector@berkeley.edu

Amy Ambrose, Sr. Development Director, amyambrose@berkeley.edu

SIMONS INSTITUTE for the Theory of Computing