

# Vinith Menon Suriyakumar

---

LAST UPDATED December 2025

CONTACT  
INFORMATION

*Email:* vinithms@mit.edu  
*Website:* VMS-6511.github.io

RESEARCH  
INTERESTS

**Areas:** Machine Learning, Statistical Inference  
**Topics:** Privacy, Security, Safety  
**Applications:** Healthcare and Law

EDUCATION

**Massachusetts Institute of Technology**, Boston, Massachusetts, USA

Department of Electrical Engineering and Computer Science  
Ph.D., Computer Science, September 2021 - Present  
Advisors: Dr. Marzyeh Ghassemi and Dr. Ashia Wilson  
Collaborators: Dr. Dylan Hadfield-Menell, Dr. Jas Sekhon  
Affiliations: LIDS, IMES, Bridgewater Associates

**University of Toronto**, Toronto, Ontario, Canada

Department of Computer Science  
M.S., Computer Science (Machine Learning), Sept 2019 - June 2021  
Focus: Differential Privacy and Algorithmic Fairness in Machine Learning for Healthcare  
Advisors: Dr. Marzyeh Ghassemi, Dr. Nicolas Papernot, Dr. Anna Goldenberg, Dr. Berk Ustun  
Affiliations: Vector Institute, The Hospital for Sick Children

**Queen's University**, Kingston, Ontario, Canada

School of Computing  
B.Computing., Biomedical Computing, May, 2019  
Thesis: Deep Classification and Generative Models for Prostate Cancer MRIs  
Advisors: Dr. Gabor Fichtinger & Dr. Parvin Mousavi  
Affiliations: Kingston Health Sciences Centre

HONORS AND  
AWARDS

NeurIPS 2025 Spotlight, September 2025  
Bridgewater Associates Research Fellowship, July 2025  
FAccT 2024 Best Paper Award, June 2024  
ICLR 2024 Oral, April 2024  
Top Reviewer, NeurIPS 2023, November 2023  
MIT Open Data Prize, MIT, October 2023  
ICML 2023 Oral, July 2023  
Wellcome Trust Fellowship, MIT, September 2021, September 2022  
ICML 2022 Spotlight, July 2022

Ethics of AI Graduate Research Fellowship, University of Toronto, August 2020

Vector Institute Research Grant, April 2020

Mitacs Accelerate Research Fellowship, December 2019

University of Toronto Arts and Science Fellowship, September 2019

Queen's University: Graduated Dean's Honor List with Distinction, June 2019

NSERC Industrial Undergraduate Research Award, August 2017

1st Degree Black Belt in Karate, October 2010

#### PUBLICATIONS

Shaib, C.\*, Suriyakumar, V.M.\*<sup>1</sup>, L.Segun, B. Wallace, M. Ghassemi. 2025. Learning the Wrong Lessons: Syntactic-Domain Shortcuts in Language Models. NeurIPS 2025. (**Spotlight, Top 3% of submissions**) \* denotes equal contribution

Suriyakumar, V.M., A. Zink, M. Hightower, M. Ghassemi, B. Beaulieu-Jones. 2025. Computational challenges arising in algorithmic fairness and health equity with generative AI. Nature Computational Science.

Suriyakumar, V.M., P. Menell, D.H. Menell, A. Wilson. The Revealed Preferences of Pre-authorized Licenses and Their Ethical Implications for Generative Models. GenLaw Workshop at ICML 2024.

Andrew, G., P. Kairouz, S. Oh, A. Opera, H. B. McMahan, V.M. Suriyakumar. 2023. One-Shot Empirical Privacy Estimation for Federated Learning. ICLR 2024. (**Oral, Top 1% of submissions**)

Jain, S., V.M. Suriyakumar, K. Creel, A. Wilson. 2023. Algorithmic Pluralism: A Structural Approach Towards Equal Opportunity. FAccT 2024. (**Best Paper Award**)

Suriyakumar, V.M., M. Ghassemi\*, B. Ustun\*. 2022. When Personalization Harms: Reconsidering the Use of Group Attributes in Prediction. ICML 2023. (**Oral, Top 1% of submissions**) \* denotes equal supervision

Suriyakumar, V.M., A. Wilson. 2022. Algorithms that Approximate Data Removal: New Results and Limitations. NeurIPS 2022.

Dziedzic, A., C.A. Choquette-Choo\*, N.Dullerud\*, V.M. Suriyakumar\*, A.S. Shamsabadi, N. Papernot, S. Jha, X. Wang. 2021. Private Multi-Winner Voting for Machine Learning. PETS 2023. \* denotes equal contribution

Amid, E., A. Ganesh, R. Matthews, S. Ramaswamy, S. Song, T. Steinke, V.M. Suriyakumar, O. Thakkar, A. Thakurta. 2021. Public Data-Assisted Mirror Descent for Private Model Training, ICML 2022 (**Spotlight, Top 3% of submissions**). (alphabetical order)

Suriyakumar, V.M., N. Papernot, A. Goldenberg, and M. Ghassemi. 2020. Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings, ACM FAccT 2021.

Cheng, V., V.M. Suriyakumar, N. Dullerud, S. Joshi, and M. Ghassemi. 2020. Can You Fake It Until You Make It?: Impacts of Differentially Private Synthetic Data on Downstream Classification Fairness, ACM FAccT 2021.

Chang, A\*, V.M. Suriyakumar\*, A. Moturu\*, N. Tewattanarat, A. Doria, and A. Goldenberg. 2020. Using Generative Models for Pediatric wbMRI. Medical Imaging in Deep Learning 2020. \* denotes equal contribution.

	<u>Suriyakumar, V.M., R. Xu, C. Pinter, G. Fichtinger. Open-source software for collision detection in external beam radiation therapy. 2017. SPIE: Journal of Medical Imaging 2017, 10135-51.</u>
IN SUBMISSION	<p>Q. Perian, <u>V.M. Suriyakumar</u>, M. Ghassemi. 2025. Iterative Nullification Transforms for Debiasing Vision-Language Models at Test-Time. In Submission.</p> <p>Xiao, Y., S. Tonekaboni, W. Gerych, <u>V.M. Suriyakumar</u>, M. Ghassemi. 2025. Mitigating Inflated Safety Risks in Language Models from Superficial Style Alignment. In Submission.</p> <p>Jin, A., W. Gerych, A. Gourabathina, <u>V.M. Suriyakumar</u>, M.Ghassemi. 2025. TOGA: Trigger Optimization for Clean Data Ordering Backdoor Attack. In Submission.</p> <p><u>Suriyakumar, V.M.</u>, A. Sekhari*, A. Wilson*. 2025. UCD: Unlearning in LLMs via Contrastive Decoding. In Submission. * denotes equal supervision</p> <p>Qian, T., <u>V.M. Suriyakumar</u>, A. Wilson, D. Hadfield-Menell. 2025. Layered Unlearning for Adversarial Relearning. In Submission.</p> <p><u>Suriyakumar, V.M.</u>, R. Alur, A. Sekhari, M. Raghavan, A. Wilson. 2024. Unstable Unlearning: The Hidden Risk of Concept Resurgence in Diffusion Models. ICLR 2025 Workshop on Navigating and Addressing Data Problems for Foundation Models.</p>
PATENTS	Leveraging Public Data in Training Neural Networks with Private Mirror Descent. U.S. Patent 20230103911. Om Dipakbhai Thakkar, Ehsan Amid, Arun Ganesh, Rajiv Mathews, Swaroop Ramaswamy, Shuang Song, Thomas Steinke, Vinith Suriyakumar, Abhradeep Guha Thakurta
	Adaptive Query Optimization Using Machine Learning. U.S. Patent 20200409948. Vincent Corvinelli, Calisto Zuzarte, Vinith Suriyakumar, Joel Raymond Scarfone, Diana Koval. <i>Patent pending.</i>
PREPRINTS	Hulkund, N.*, <u>V.M. Suriyakumar</u> , T. Killian, M. Ghassemi. 2022. Improving Robustness to Distribution Shift with Algorithmic Stability. (in submission) * denotes equal contribution
	Chang, A., <u>V.M. Suriyakumar</u> *, A. Moturu*, J. Tu, N. Tewattanarat, S. Joshi, A. Doria, and A. Goldenberg. 2021. Incorporating 3D Context to Unsupervised Cancer Detection in Pediatric WbMRI. * denotes equal contribution
BOOKS & CHAPTERS	Differential Privacy and Medical Data Analysis, Differential Privacy for Artificial Intelligence Applications. Now Publisher Inc. <u>Suriyakumar, V.M.</u> , N. Papernot, and A. Goldenberg. 2024. (Forthcoming)
INVITED TALKS	<p><u>Safely Open-Sourcing Foundation Models</u>, MITAI Conference, October 2025</p> <p><u>Just Forget About It: Lessons from the Frontiers of Machine Unlearning</u>, Bridgwater AIA Labs Distinguished Speaker Series, August 2025</p> <p><u>Unlearning &amp; Privacy</u>, 6.3950 AI, Decision Making and Society, MIT, November 2024</p> <p><u>Tradeoffs in Trustworthy Machine Learning</u>, 6.S977 Ethical Machine Learning in Human Deployments , MIT, March 2024</p> <p><u>Personalization Harms: Reconsidering the Use of Group Attributes in Prediction</u>, Collaborative Data Science for Healthcare, Harvard T.H. Chan School of Public Health, August 2023</p>

Tradeoffs in Privacy, Utility, and Fairness, 6.S977 Ethical Machine Learning in Human Deployments  
, MIT, March 2022

Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings, Ethics of AI in Context: Emerging Scholars, Centre for Ethics, University of Toronto, October 2020

Chasing Your Long Tails: Differentially Private Prediction in Health Care Settings, Vector Institute, University of Toronto, October 2020

**SKILLS**  
Data Processing Frameworks: Pandas, Numpy  
Machine Learning Frameworks: transformers, diffusers, Tensorflow, PyTorch  
ML DevOps Frameworks: Weights and Biases, Tensorboard  
Languages: Python, R

**MENTORING AND ADVISING**  
Julia Liu, Undergraduate Researcher, Fall 2025 - Present  
Timothy Qian, M.Eng, Fall 2024 - Spring 2025  
Helen Propson, Research Fellow, Spring 2024 - Spring 2025  
Quinn Perian, Undergraduate Researcher, Spring 2024 - Spring 2025  
Neha Hulkund, Undergraduate Researcher, Spring 2021 - Spring 2022  
Shrey Jain, Undergraduate Researcher, Summer 2020 - Winter 2021  
Victoria Cheng, Undergraduate Researcher, Summer 2020

**REVIEWING AND ORGANIZING**  
Program Committee, IJCAI 2020 AI for Social Good Workshop  
Program Committee, NeurIPS 2020 Machine Learning for Health Workshop  
External Reviewer, USENIX Security 2021  
Reviewer, Journal of Artificial Intelligence Research, 2021  
External Reviewer, ICML 2021  
External Reviewer, NeurIPS 2021  
Reviewer, NeurIPS 2021 Datasets and Benchmarks Track  
Reviewer, Journal of Privacy and Confidentiality  
Program Committee, FAccT 2022  
Reviewer, ICML 2022  
Organizer, NeurIPS 2022 Robustness in Sequence Modelling Workshop  
Program Committee, IEEE SaTML 2023  
Reviewer, NeurIPS 2023  
Program Committee, FAccT 2024  
Reviewer, ICML 2024  
Reviewer, ICLR 2025  
Reviewer, ICML 2025  
Area Chair, NeurIPS 2025

**SELECTED PROFESSIONAL EXPERIENCE**  
**Bridgewater Associates**, Remote  
**Research Fellow** **June 2025 - Present**  
I collaborate with the folks at Bridgewater's AIA lab led by Dr. Jas Sekhon on intepretability and safety concerns regarding LLMs.

**Mass General Brigham**, Remote

***Research Trainee***

**January 2024 - May 2024**

I worked with Dr. David Bates to develop models for predicting adverse pregnancy outcomes using pre-pregnancy data.

**Kaiser Permanente**, Remote

***Graduate Research Affiliate***

**September 2023 - May 2024**

I worked with Dr. Yeyi Zhu to develop models for predicting adverse pregnancy outcomes using pre-pregnancy data.

**Google**, Remote

***Student Researcher***

**May 2022 - August 2022**

I worked with Dr. Galen Andrew, Dr. Peter Kairouz, and Dr. Sewoong Oh on developing methods to empirically measure the privacy leakage (specifically differential privacy) of federated learning mechanisms.

**Google**, Remote

***Research Intern***

**May 2021 - August 2021**

Building new algorithms to improve the utility of differentially private federated learning using public data with Dr. Om Thakkar, Swaroop Ramaswamy and collaborators at Google Brain Privacy and Security.

**The Hospital for Sick Children**, Toronto, Ontario Canada

***Research Assistant***

**May, 2019 - May 2021**

Building anomaly detection methods using generative models for early detection of pediatric cancer in whole body MRIs. This project is in collaboration with clinicians in the SickKids' Radiology department.

**Cape Privacy (formerly Dropout Labs)**, Remote

***Consultant***

**June, 2019 - August, 2019**

Contributed tutorials to the open-source library TF Encrypted for machine learning under secure multiparty computation protocols. Started investigations into using self-learning activation functions using polynomial approximations to speed up training time.

**Square**, San Francisco, California USA

***Data Science Intern***

**May, 2018 - August, 2018**

Developed a representation learning algorithm to cluster merchants into different business categories for improved pricing algorithms with 90% accuracy. Involved in ethics and governance of AI in products committee analyzing what Square's principles would be when implementing AI into its products.

**Helpful (acquired by Shopify)**, Toronto, Ontario Canada

***Machine Intelligence Intern***

**September, 2017 - April, 2018**

Improved transcriptions for speech recognition problems such as getting names and company specific jargon correct by 4-10x. Investigated computational linguistic techniques such as phoneme matching and pronunciation modelling to further improve transcriptions in the presence of different accents.

**IBM**, Toronto, Ontario Canada

***Deep Learning and Systems Research Intern***

**May, 2017 - August, 2017**

Led a research project exploring improvements to traditional query optimization in databases using machine learning. Implemented a few shot learning algorithm based on matching networks improving

the database speed by 30% across standard SQL query speed benchmarks. Currently, I have 1 patent pending from this work.

## SERVICE AND VOLUNTEERING

VP Visit Days & Orientation, MIT EECS Graduate Student Association January  
2022 - January 2023

I led a team of 12 students to redesign the Visit Days and Orientation for prospective and incoming PhD students for the MIT EECS department. These are three-day and week-long events, respectively, that are essential to helping new students feel welcome. This involved large-scale planning and execution over the course of the year managing a \$50,000 budget.

**Director of Finance & Advisor, CUSEC**      **January 2019 - February 2021**

I manage a budget of approximately \$100,000 for a nationwide software engineering conference of 500 students. The conference brings over 15 industry sponsors and 20 speakers from all over North America to Montreal for three days to engage in a variety of topics in software engineering. I advise the chairs and the conference organizers on best practices.

NeurIPS 2019 Student Volunteer December 2019

I was selected to be a student volunteer in helping run this premier machine learning research conference that brings over 12 000 researchers from all over the world. My role involved organizing attendees into different lectures and paper presentations.

Co-Chair, Toronto Health Data Hackathon      September 2019 - October 2019

I led a team of 5 to organize this important hackathon in collaboration with the Vector Institute and St. Michael's Hospital. The event gathered 100 computer scientists and doctors to build new machine learning for health products over the course of two days.

Chair, QHacks April 2018 - February 2019

I lead a team of 17 students to create a 500 person hackathon to engage and empower students to build products and connect with the tech industry. I developed a sustainable internal operating structure focusing on team autonomy and transparency. Provided bi-weekly mentorship to each individual to ensure important growth in desired areas. I ran discussions on gender and racial discrimination in tech and how we as an organization can support these marginalized groups.

Co-Chair, CUSEC January 2018 - January 2019

I lead a team of 25 students remotely to create a 500 person conference to engage and empower students to explore different areas of the software engineering industry. I improved engagement across a number of Canadian universities and engaged a more diverse set of speakers so gender, racial, and sexual orientation representation were present. Provided bi-weekly mentorship to each individual to ensure important growth in desired areas.

**Director of Events, CUSEC**      **January 2017 - January 2018**

I managed and executed the logistics for five different events and 12 different workshops at the scale of 500 attendees. Led the pilot of a new event to increase engagement between students about pressing issues of gender and racial discrimination.

VP Operations, Queen's Computing Students' Association March 2017 - April 2018

I hired, led, and supported a team of 7 commissioners who lead efforts in academics, casual events, formal events, marketing, finance, equity and governance. Restructured our hiring process to reduce biases and improve equity. I piloted a first year internship program within the association which increased first year student engagement by 50%.