Charith Mendis

Assistant Professor, Siebel School of Computing and Data Science

University of Illinois at Urbana-Champaign
4118, *Siebel School of Computing and Data Science*
201 *N. Goodwin Ave., Urbana, IL* 61801, *USA*
+1 (917) 780 9232
charithm@illinois.edu
http://charithmendis.com/

Professional Experience

UIUC **Assistant Professor**, *Siebel School of Computing and Data Science, University of Illinois at Urbana-Champaign*, Urbana, IL (Aug. 2021 - Present)

Amazon **Amazon Scholar**, *Annapurna Labs, Amazon*, Virtual (Mar. 2024 - Present)

Google **Visiting Faculty Researcher**, *Google Brain*, New York, NY (Aug. 2020 - Dec. 2021)

Microsoft **Research Intern**, *Microsoft Research*, Redmond, WA (Jun. 2015 - Sep. 2015)

Education

2013–2020 **Ph.D. in Computer Science**, *Massachusetts Institute of Technology*, Cambridge, MA
- Thesis: Towards Automated Construction of Compiler Optimizations
- Advisor: Saman Amarasinghe

2013–2015 **S.M. in Computer Science**, *Massachusetts Institute of Technology*, Cambridge, MA
- Thesis: Helium: Lifting High-Performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code
- Advisor: Saman Amarasinghe
- Awards: **William A. Martin Memorial Thesis Award for outstanding SM thesis in CS**

2008–2013 **B.Sc. in Electronics and Telecommunication Engineering**, *University of Moratuwa*, Sri Lanka
- Awards: Ranked 1st out of 650 students (Gold Medalist).

Research Awards and Honors

2026	DARPA YFA Director's Fellowship (outstanding YFA performance)	*DARPA*
2026	ISPASS Honorable Mention for the Distinguished Artifact Award	*ISPASS*
2025	Google ML and Systems Junior Faculty Award	*Google*
2025	SIGMOD Honorable Mention for the Best Artifact Award	*SIGMOD*
2025	POPL Distinguished Paper Award	*POPL*
2024	DARPA Young Faculty Award (YFA)	*DARPA*
2024	NSF CAREER Award	*NSF*
2024	IEEE Micro Top Picks Honorable Mention	*IEEE Micro*
2019	ML for Systems Workshop Best Paper Award	*ML for Systems Workshop @ISCA 2019*
2017	IEEE BigData Best Student Paper Award	*IEEE International Conference on Big Data (BigData)*
2015	William A. Martin Memorial Thesis Award for an outstanding SM thesis in CS	*MIT*
2013	MIT Energy Initiative Fellowship	*MIT*
2012	Sri Lanka Telecom Gold Medal for the Best Academic Performance	*University of Moratuwa, Sri Lanka*

Teaching and Service Awards and Honors

2024	Outstanding Advisor Award, Grainger College of Engineering	*UIUC*
Fall 2023	List of Teachers Ranked as Excellent by Their Students (CS 598 CM)	*UIUC*
Fall 2022	List of Teachers Ranked as Excellent by Their Students (CS 598 CM)	*UIUC*
Fall 2021	List of Teachers Ranked as Excellent by Their Students (CS 598 CM)	*UIUC*
2021	Outstanding Reviewer	*ICLR*

Research Grants

*I have raised more than **$4.7M** in funding as a PI or a co-PI (more than **$13.7M** to the university in total) through research grants and industrial gifts. These include **NSF CAREER, DARPA YFA, and the Google ML and Systems Junior Faculty Awards.***

2025-2028 **DARPA MOCHA (UIUC co-PI)**, "Compiler 2.0: Revisiting the Construction of a Compiler", PI: Michael Gordon (Aarno Labs), co-PIs: Edgar Solomonik (UIUC) and 5 co-PIs from MIT, I am leading this grant from the UIUC side.
UIUC Total: $931,254, My Share: $660,735
Dates: 09/05/2025 - 09/04/2028

2025-2027 **IBM Illinois Discovery Accelerator Institute (UIUC PI)**, "Algorithmic, systems and compiler efficiency improvements for optimizing deep neural networks", co-PIs: Deming Chen and Sasa Misailovic
UIUC Total: $689,236, My Share: $240,184
Dates: 08/16/2025 - 08/15/2027

2025-2025 **IBM Illinois Discovery Accelerator Institute (UIUC PI)**, "Optimized GPU Code Generation Framework for Sparse Regular Attention"
UIUC Total: $60,000, My Share: $60,000
Dates: 01/01/2025 - 08/15/2025

2024-2027 **DARPA YFA (UIUC PI)**, "Language and Hardware Adaptive Representations and Techniques for Compiling Heterogeneous Workloads"
UIUC Total: $1M, My Share: $1M
Dates: 09/01/2024 - 08/31/2027
Awarded the DARPA Director's Fellowship.

2024-2029 **NSF CAREER (UIUC PI)**, "CAREER: An Agile Compiler Framework for Spatial Dataflow Accelerators"
UIUC Total: $525,000, My Share: $525,000
Dates: 02/01/2024 - 01/31/2029

2023-2028 **NSF PPoSS (UIUC co-PI)**, *PPoSS Large grant*, "Collaborative Research: PPoSS: LARGE: General-Purpose Scalable Technologies for Fundamental Graph Problems", lead PI: Josep Torrellas, altogether 10 PIs from UIUC, MIT, and Indiana University
UIUC Total: $3.9M My Share: $728,000
Dates: 08/01/2023 - 07/31/2028

2023-2027 **SRC JUMP 2.0 (UIUC co-PI)**, *JUMP* 2.0 *Center sponsored partly by DARPA*, "ACE: Evolvable Computing for Next Generation Distributed Computer Systems", Director and lead PI: Josep Torrellas, altogether 21 PIs from UIUC, Cornell, Georgia Tech, MIT, Ohio State, Purdue University, Stanford, the University of California San Diego, the University of Kansas, the University of Michigan, the University of Texas at Austin, and the University of Washington with total funding of $39.6M
UIUC total: $6.27M, My Share: $1.188M
Dates: 01/01/2023 - 12/31/2027

Industry Funding

Unrestricted Gifts
- **Google (2025):** for the ML and Systems Junior Faculty Award (total: $100,000)
- **Qualcomm (2024):** for exploring novel auto-tuning techniques (total: $100,000)
- **Qualcomm (2023):** for research on a synthesis-based retargetable compiler (total: $106,000 shared with Vikram Adve; my share: $53,000)
- **Intel (2022):** for research on automatically generating MLIR code (total: $70,000 shared with Vikram Adve; my share: $35,000)
- **Google (2021):** for research on ML for compilers and compilers for ML (total: $30,000)

Cloud Credits
- **AWS Cloud Credits (2025)**: Build on Trainium (total: $50,000)
- **Google Cloud Credits (2023)**: for sparse transformer research (total: $53,700)
- **AWS Cloud Credits (2022)** for graph neural network research (total: $20,620)

Journal and Conference Publications

*Published 40 Journal/Conference papers (**30 papers after joining Illinois**) as of April 15, 2026. Up-to-date citation metrics can be found in my Google Scholar page (`https://scholar.google.com/citations?user=utZ3JYUAAAAJ&hl=en`). My name is underlined, while my students' names are bolded. * means equal contribution.*

OSDI 2026. **Muyan Hu**, **Ahan Gupta**, Jiachen Yuan, Vima Gupta, **Taeksang Kim**, Xin Ju, Janardhan Kulkarni, Ofer Dekel, Vikram Adve, and Charith Mendis. VTC: DNN Compilation with Virtual Tensors for Data Movement Elimination. *In 20th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, 2026.

EuroSys 2026. Yuan Zhou, Shaojie Xiang, Lingfan Yu, Zhenyu Song, Charith Mendis, and Yida Wang. SAS: Sparse Attention Synthesizer for Efficient Language Model Inference. *In European Conference on Computer Systems (EuroSys)*, 2026.

ISPASS 2026. Jianming Tong, Yujie Li, **Devansh Jain**, Charith Mendis, and Tushar Krishna. MINISA: Minimal Instruction Set Architecture for Next-gen Reconfigurable Inference Accelerator. *In IEEE International Symposium on Performance Analysis of Systems and Software (ISPASS)*, 2026 [**Honorable Mention for the Distinguished Artifact Award**].

CGO 2026. **Damitha Lenadora**, **Vimarsh Sathia**, Gerasimos Gerogiannis, Serif Yesil, Josep Torrellas, and Charith Mendis. GRANII : Selection and Ordering of Primitives in GRAph Neural Networks using Input Inspection. *In International Symposium on Code Generation and Optimization (CGO)*, 2026.

MICRO 2025. **Devansh Jain**, **Marco Frigo**, **Jai Arora**, **Akash Pardeshi**, **Zhihao Wang**, **Krut Patel**, and Charith Mendis. TAIDL: Tensor Accelerator ISA Definition Language with Auto-generation of Scalable Test Oracles. *In 58th IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2025 [**Adopted by Amazon Trainium Accelerators**].

POPL 2025. **Jai Arora**, Sirui Lu, **Devansh Jain**, **Tianfan Xu**, Farzin Houshmand, Phitchaya Mangpo Phothilimthana, Mohsen Lesani, Praveen Narayanan, Karthik Srinivasa Murthy, Rastislav Bodik, Amit Sabne, and Charith Mendis. TensorRight: Automated Verification of Tensor Graph Rewrites. *Proc. ACM Program. Lang. 9, POPL, Article 29*, 2025 [**Distinguished Paper Award**].

PLDI 2025. Abdul Rafae Noor, Dhruv Baronia, Akash Kothari, Muchen Xu, Charith Mendis, and Vikram Adve. MISAAL: Synthesis-Based Automatic Generation of Efficient and Retargetable Semantics-Driven Optimizations. *Proc. ACM Program. Lang. 9, PLDI, Article 198*, 2025.

OOPSLA 2025. **Damitha Lenadora**, **Nikhil Jayakumar**, **Chamika Sudusinghe**, and Charith Mendis. GALA: A High Performance Graph Neural Network Acceleration LAnguage and Compiler. *Proc. ACM Program. Lang. 9, OOPSLA2, Article 335*, 2025.

OOPSLA 2025. **Avaljot Singh**, Yasmin Chandini Sarita, Charith Mendis, and Gagandeep Singh. Automated Verification of Soundness of DNN Certifiers. *Proc. ACM Program. Lang. 9, OOPSLA1, Article 144*, 2025.

OOPSLA 2025. **Ahan Gupta**, **Yueming Yuan**, **Devansh Jain**, **Yuhao Ge**, David Aponte, Yanqi Zhou, and Charith Mendis. SPLAT: A Framework for Optimised GPU Code-Generation for SParse reguLar ATtention. *Proc. ACM Program. Lang. 9, OOPSLA1, Article 138*, 2025.

ICML 2025. **Chamika Sudusinghe**, Gerasimos Gerogiannis, **Damitha Lenadora**, Charles Block, Josep Torrellas, and Charith Mendis. COGNATE: Acceleration of Sparse Tensor Programs on Emerging Hardware using Transfer Learning. *In Forty-second International Conference on Machine Learning (ICML)*, 2025.

SIGMOD 2025. Yuxuan Zhu, Tengjun Jin, **Stefanos Baziotis**, Chengsong Zhang, Charith Mendis, and Daniel Kang. PilotDB: Database-Agnostic Online Approximate Query Processing with A Priori Error Guarantees. *In Proceedings of ACM on Management of Data (SIGMOD)*, 2025.

SIGMOD 2024. **Stefanos Baziotis**, Daniel Kang and Charith Mendis. Dias: Dynamic Rewriting of Pandas Code. *In Proceedings of ACM on Management of Data (SIGMOD)*, 2024 [**Honorable Mention for the Best Artifact Award**].

ASPLOS 2024. **Yufeng Wang** and Charith Mendis. TGLite: A Lightweight Programming Framework for Continuous-Time Temporal Graph Neural Networks. *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2024.

ASPLOS 2024. Charles Block, Gerasimos Gerogiannis, Charith Mendis, Ariful Azad and Josep Torrellas. Two-Face: Combining Collective and One-Sided Accesses for Efficient Distributed SpMM. *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2024.

ASPLOS 2024. Akash Kothari, Abdul Rafae Noor, Muchen Xu, Hassam Uddin, Dhruv Baronia, **Stefanos Baziotis**, Vikram Adve, Charith Mendis and Sudipta Sengupta. Hydride: A Retargetable and Extensible Synthesis-based Compiler for Modern Hardware Architectures. *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2024.

SAS 2024. **Avaljot Singh**, Yasmin Sarita, Charith Mendis, and Gagandeep Singh. ConstraintFlow: A Declarative DSL for Easy Development of DNN Certifiers. *In Static Analysis: 31st International Symposium (SAS)*, 2024.

CIKM 2024 Ruijie Wang, **Wanyu Zhao**, Dachun Sun, Charith Mendis, and Tarek Abdelzaher. Towards Efficient Temporal Graph Learning: Algorithms, Frameworks, and Tools. *In Proceedings of the 33rd ACM International Conference on Information and Knowledge Management (CIKM)*, 2024.

SIGIR 2024. Ruijie Wang, Jingyuan Huang, Yutong Zhang, Jinyang Li, **Yufeng Wang**, **Wanyu Zhao**, Shengzhong Liu, Charith Mendis and Tarek Abdelzaher. TGOnline: Enhancing Temporal Graph Learning with Adaptive Online Meta-Learning. *Proceedings of the 47th International Conference on Research and Development in Information Retrieval (SIGIR)*, 2024.

MLSys 2024. Isha Chaudhary, Alex Renda, Charith Mendis, and Gagandeep Singh. COMET: Neural Cost Model Explanation Framework. *In 7^{th} Annual Conference on Machine Learning and Systems (MLSys)*, 2024.

PPoPP 2023. **Yufeng Wang** and Charith Mendis. TGOpt: Redundancy-Aware Optimizations For Temporal Graph Attention Networks. *International Symposium on Principles and Practice of Parallel Programming (PPoPP)*, 2023.

ISCA 2023. Gerasimos Gerogiannis, Serif Yesil, **Damitha Lenadora**, Dingyuan Cao, Charith Mendis and Josep Torrellas. SPADE: A Flexible and Scalable Accelerator for SpMM and SDDMM. *International Symposium on Computer Architecture (ISCA)*, 2023 [**IEEE Micro Top Picks Honorable Mention**].

NeurIPS 2023. Phitchaya Mangpo Phothilimthana, Sami Abu-El-Haija, Kaidi Cao, Bahare Fatemi, Charith Mendis, and Bryan Perozzi. TpuGraphs: A Performance Prediction Dataset on Large Tensor Computational Graphs. *37^{th} Annual Conference on Neural Information Processing Systems (NeurIPS): Benchmarks and Dataset Track*, 2023 [**Real world dataset of large computational graphs**].

NeurIPS 2023. Kaidi Cao, Phitchaya Mangpo Phothilimthana, Sami Abu-El-Haija, Dustin Zelle, Yanqi Zhou, Charith Mendis, Jure Leskovec and Bryan Perozzi. Learning Large Graph Property Prediction via Graph Segment Training. *37th Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2023.

MetaCom 2023. Tarek Abdelzaher, Matthew Caesar, Charith Mendis, Klara Nahrstedt, Mani Srivastava and Minlan Yu. Challenges in Metaverse Research: An Internet of Things Perspective. *IEEE International Conference on Metaverse Computing, Networking and Applications*, 2023 **[Invited Paper]**.

MLSys 2023. Jaeyeon Won, Changwan Hong, Charith Mendis, Joel Emer and Saman Amarasinghe. Unified Convolution Framework: A compiler-based approach to support sparse convolutions. *In 6th Annual Conference on Machine Learning and Systems (MLSys)*, 2023.

ASPLOS 2023. Jaeyeon Won, Charith Mendis, Joel Emer and Saman Amarasinghe. WACO: Learning workload-aware co-optimization of the format and schedule of a sparse tensor program. *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2023.

IISWC 2022. Ondrej Sykora, Mangpo Phothilimthana, Charith Mendis and Amir Yazdanbakhsh. GRANITE: A Graph Neural Network Model for Basic Block Throughput Estimation. *IEEE International Symposium on Workload Characterization (IISWC)*, 2022.

PLDI 2022. Yishen Chen, Charith Mendis and Saman Amarasinghe. All you need is SLP: Systematic Control-Flow Vectorization with SLP. *Proceedings of the 43rd ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, 2022.

ASPLOS 2021. Yishen Chen, Charith Mendis, Michael Carbin and Saman Amarasinghe. Vegen: A Vectorizer Generator for SIMD and Beyond. *International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, 2021.

MLSys 2021. Samuel Kaufman, Phitchaya Phothilimthana, Yanqi Zhou, Charith Mendis, Sudip Roy, Amit Sabne and Mike Burrows. A Learned Performance Model for Tensor Processing Units. *In 4th Annual Conference on Machine Learning and Systems (MLSys)*, 2021 **[Productionized at Google within the XLA TPU compiler, featured in the Google Research blog]**.

Before joining UIUC (during PhD):

MICRO 2020. Alex Renda, Yishen Chen, Charith Mendis and Michael Carbin. Optimizing CPU Simulator Parameters with Learned Differentiable Approximations. *In 53rd IEEE/ACM International Symposium on Microarchitecture (MICRO)*, 2020.

NeurIPS 2019. Charith Mendis, Cambridge Yang, Yewen Pu, Saman Amarasinghe and Michael Carbin. Compiler Auto-Vectorization with Imitation Learning. *32nd Conference on Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

IISWC 2019. Yishen Chen, Ajay Brahmakshatriya, Charith Mendis, Alex Renda, Eric Atkinson, Ondrej Sykora, Saman Amarasinghe and Michael Carbin. BHive: A Benchmark Suite and Measurement Framework for Validating x86-64 Basic Block Performance Models. *IEEE International Symposium on Workload Characterization (IISWC)*, 2019.

ICML 2019. Charith Mendis, Alex Renda, Saman Amarasinghe and Michael Carbin. Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks. *36th International Conference on Machine Learning (ICML)*, 2019 **[Best Paper Award - ML for Systems Workshop @ISCA 2019]**.

CC 2019. Charith Mendis*, Ajay Jain*, Paras Jain and Saman Amarasinghe. Revec: Program Rejuvenation through Revectorization. *Proceedings of the 28th International Conference on Compiler Construction (CC)*, 2019.

OOPSLA 2018. Charith Mendis and Saman Amarasinghe. goSLP: Globally Optimized Superword Level Parallelism Framework. *Proc. ACM Program. Lang. 2, OOPSLA*, 2018.

BigData 2017. Yunming Zhang, Vladimir Kiriansky, Charith Mendis, Matei Zaharia and Saman Amarasinghe. Making Caches Work for Graph Analytics. *IEEE International Conference on Big Data (BigData)*, 2017 [**Best Student Paper Award**].

ICASSP 2016. Charith Mendis, Jasha Droppo, Saeed Maleki, Madanlal Musuvathi, Todd Mytkowicz and Geoffrey Zweig. Parallelizing WFST Speech Decoders. *IEEE International Conference on Acoustics, Speech and Signal Processing, (ICASSP)*, 2016.

PLDI 2015. Charith Mendis, Jeffrey Bosboom, Kevin Wu, Shoaib Kamil, Jonathan Ragan-Kelley, Sylvain Paris, Qin Zhao and Saman Amarasinghe. Helium: lifting high-performance stencil kernels from stripped x86 binaries to Halide DSL code. *Proceedings of the 36th ACM SIGPLAN Conference on Programming Language Design and Implementation (PLDI)*, 2015 [**Featured in Fortune News**].

Theses

Charith Mendis. Towards Automated Construction of Compiler Optimizations. *PhD Thesis, Massachusetts Institute of Technology*, 2020.

Charith Mendis. Helium: Lifting High-Performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code. *SM Thesis, Massachusetts Institute of Technology*, 2015 [**William A. Martin Memorial Thesis Award**].

Workshop Papers

LCPC 2023 Pavlo Pastaryev, Charith Mendis, and Lawrence Rauchwerger. Generating memory allocators from the ground up. *36th International Workshop on Languages and Compilers for Parallel Computing*, 2023.

EXAIT@ICML 2025 **Chamika Sudusinghe**, Gerasimos Gerogiannis, **Damitha Lenadora**, Charles Block, Josep Torrellas, Charith Mendis. Automated data selection for efficient cost model training to optimize sparse matrix kernels on emerging hardware accelerators. *Exploration in AI Today (EXAIT) @ ICML'25*, 2025.

Books Edited

LNCS 2023 Charith Mendis, and Lawrence Rauchwerger. Languages and Compilers for Parallel Computing: 35th International Workshop, LCPC 2022, Chicago, IL, USA, October 12–14, 2022, Revised Selected Papers. *Lecture Notes in Computer Science (LNCS), Volume 13829*, 2023.

Tutorials

Devansh Jain, **Akash Pardeshi**, **Marco Frigo**, **Kaustubh Khulbe** and Charith Mendis. End-to-End Compiler Infrastructure for Emerging Tensor Accelerators. *2025-present*, (OOPSLA 2025, MICRO 2025, ASPLOS 2026, PLDI 2026).

Jianming Tong, Yujie Li, Tushar Krishna, Zhiru Zhang, Hongzheng Chen, Niansong Zhang, **Akash Pardeshi**, **Devansh Jain** and Charith Mendis. RAIC: Reconfigurable AI Computing. *2026-present*, (ASPLOS 2026).

Preprints

- **Devansh Jain**, **Akash Pardeshi**, **Marco Frigo**, **Krut Patel**, **Kaustubh Khulbe**, **Jai Arora**, and Charith Mendis. ACT: Automatically Generating Compiler Backends from Tensor Accelerator ISA Descriptions. ArXiv:`https://arxiv.org/pdf/2510.09932`

- **Avaljot Singh**, Yamin Chandini Sarita, Aditya Mishra, **Ishaan Goyal**, Gagandeep Singh, and Charith Mendis. A Tensor-Based Compiler and a Runtime for Neuron-Level DNN Certifier Specifications. ArXiv:`https://arxiv.org/abs/2507.20055`
- **Alex Broihier**, **Stefanos Baziotis**, Daniel Kang and Charith Mendis. PandasBench: A Benchmark for the Pandas API. ArXiv:`https://arxiv.org/abs/2506.02345`
- **Ahan Gupta**, **Yueming Yuan**, **Hao Guo**, Yanqi Zhou, and Charith Mendis. FLuRKA: Fast fused Low-Rank & Kernel Attention. ArXiv:`https://arxiv.org/abs/2306.15799`
- **Wanyu Zhao**, **Yuhao Ge**, **Chamika Sudusinghe** and Charith Mendis. TGCache: Reducing Data Movement for Large-Scale Temporal GNN Training.
- **Stefanos Baziotis** and Charith Mendis. metap: A meta-programming layer for Python.

Student Advising

I advise nine PhD students (two co-advised), six MS theses students (11 in total), and mentor 24 undergraduate students. The papers the students have published with me are briefly mentioned here.

Current PhD Students

- Damitha Lenodara (from Fall 2021; Thesis proposal defended: Jan 2026, expected graduation: 2027)
  - *Thesis Title:* Abstractions and Acceleration of Graph Neural Networks
  - GRANII [CGO'26], GALA [OOPSLA'25], COGNATE [ICML'25], SPADE [ISCA'23 **(IEEE Micro Top Picks Honorable Mention)**]
- Stefanos Baziotis (from Fall 2021; Thesis proposal defended: Feb 2026, expected graduation: 2027)
  - *Thesis Title:* Optimizing Data Analytics Using Compiler Technology
  - PilotDB [SIGMOD'25], Hydride [ASPLOS'24], Dias [SIGMOD'24 **(Honorable Mention for the Best Artifact)**]
- Avaljot Singh (co-advised with Gagandeep Singh from Fall 2022; Thesis proposal defended: Apr 2026, expected graduation: 2027)
  - *Thesis Title:* Building Abstractions for Bridging Formal Methods and Machine Learning
  - ProveSound [OOPSLA'25], ConstraintFlow [SAS'24]
- Devansh Jain (from Fall 2023; expected graduation: 2028)
  - MINISA [ISPASS'26 **(Honorable Mention for the Distinguished Artifact)**], TAIDL [MICRO'25], TensorRight [POPL'25 **(Distinguished Paper)**], SPLAT [OOPSLA'25]
- Chamika Sudusinghe (from Fall 2023; expected graduation: 2028) [**NSF GFRP Fellow, IEEE-USA George F. McClure Citation of Honor**]
  - COGNATE [ICML'25], GALA [OOPSLA'25]
- Jai Arora (from Fall 2023; expected graduation: 2028)
  - TensorRight [POPL'25 **(Distinguished Paper)**], TAIDL [MICRO'25]
- Muyan Hu (co-advised with Vikram Adve from Fall 2023; expected graduation: 2028) [**ML and Systems Rising Star Award**]
  - VTC [OSDI'26]
- Heng Zhong (from Fall 2025; expected graduation: 2030)
- Zirui Zhou (from Fall 2025; expected graduation: 2030)

Rotating PhD Students

The following includes work done while rotating in my lab.

- Ahan Gupta (Fall 2022 - Fall 2024)
  - SPLAT [OOPSLA'25], VTC [OSDI'26]
- Wanyu Zhao (Fall 2023 - Fall 2024)
  - TGOnline [SIGIR'24], TGNN-Tutorial [CIKM'24]

Graduated Masters Students

- Yufeng Wang (MS UIUC 2021-2023)
  - *Thesis Title:* A Framework for Programming and Optimizing Temporal Graph Neural Networks
  - TGOpt [PPoPP'23], TGLite [ASPLOS'24], TGOnline [SIGIR'24]
  - **David J. Kuck Outstanding Master's Thesis Award 2024**
  - *Next* → Tesla Compiler Engineer

Graduated Masters Students (cntd.)
- Hao Guo (MS UIUC 2023-2025)
  - *Thesis Title:* Ghostdecoding: leveraging random-feature kernels for error-aware and training-free KV cache selection
  - *Next* → NVidia Compiler Engineer
- Krut Patel (MS UIUC 2023-2025)
  - *Thesis Title:* Optimization Opportunities for Various Heterogeneous Pipelines
  - TAIDL [MICRO'25]
  - *Next* → NVidia Compiler Engineer
- Dhruv Sundararam (MCS 2024-2025), *Next* → NVidia
- Yuhao Ge (MCS 2023-2024), *Next* → AWS Annapurna Labs → Google
  - SPLAT [OOPSLA'25]
- Taeksang Kim (MEng 2024-2025), *Next* → Together AI
  - VTC [OSDI'26]

Current Masters Students
- Akash Pardeshi (from Fall 2024; expected graduation: 2026): MS Theses, *Next* → MIT PhD
- Saatvik Lochan (from Fall 2025; expected graduation: 2027): MS Theses
- Bhavya Hirani (from Fall 2025; expected graduation: 2027): MS Theses
- Dushyant Bharadwaj (from Fall 2025; expected graduation: 2027): MCS
- Kaushik Varadharajan (from Fall 2025; expected graduation: 2027): MCS

Undergraduate Advising
- Alan Xie (Spring 2026 - Present)
- Pedro Couto (Spring 2026 - Present)
- Ryan Chan (Spring 2026 - Present)
- Tanmay Garudadri (Spring 2026 - Present)
- Mihir Tandon **[James Scholar]** (Spring 2026 - Present)
- Jacob Xianyu (Summer 2025 - Spring 2026)
- Advait Tahilyani (Summer 2025 - Fall 2025)
- Ishaan Goyal **[James Scholar]** (Summer 2025) **[co-author of a submitted paper]**
- Kaustubh Khulbe (Fall 2024 - Fall 2025) **[co-author of a submitted paper]**
- Ethan Sirois (Fall 2024 - Spring 2025; co-founder CPA Dome)
- Rayan Singh (Fall 2024 - Spring 2025; Next: MS at UIUC)
- Max Gendeh (Fall 2024 - Spring 2025; Junior)
- Marco Frigo (Summer 2024 - Present)
- Emily Hsieh (Summer 2024 - Spring 2025; Junior)
- Zhihao Wang (Summer 2024 - Spring 2025; Next: MS at UIUC)**[completed senior thesis, co-author of MICRO'25]**
- Andy Luo (Spring 2024 - Fall 2025)
- Alex Broihier (Spring 2024 - Spring 2025; Junior) **[co-lead of a submitted paper]**
- Kaushik Varadharajan **[James Scholar]** (Spring 2024 - Spring 2025; Next: MS at UIUC)
- Nikhil Jayakumar (Spring 2024 - Present) **[co-author of OOPSLA'25]**
- Hengjia Yu (Spring 2024 - Fall 2024)
- Matthew Jacob (Fall 2023; Senior)
- Sun I (Spring 2023 - Summer 2023; Sophomore)
- Yueming Yuan (Fall 2022 - Fall 2023; Next: PhD at UIUC) **[co-author of OOPSLA'25]**
- Tianfan Xu (Summer 2022 - Summer 2023; Next: MS at Harvard) **[co-author of POPL'25]**

Teaching

Spring 2026	Instructor for Advanced Compiler Construction (CS 526)
Fall 2025	Instructor for ML and Compilers (CS 521 LCU/LCC)
Spring 2025	Instructor for ML and Compilers (CS 521 LCU/LCC)
Fall 2024	Instructor for Compiler Construction (CS 426)
Spring 2024	Instructor for Advanced Compiler Construction (CS 526)
Fall 2023	Instructor for Machine Learning for Compilers and Architecture (CS 598 CM), **Selected to the "List of Teachers Ranked as Excellent By Their Students".**

Fall 2022 Instructor for Machine Learning for Compilers and Architecture (CS 598 CM), **Selected to the "List of Teachers Ranked as Excellent By Their Students".**

Spring 2022 Instructor for Advanced Compiler Construction (CS 526)

Fall 2021 Instructor for Machine Learning for Compilers and Architecture (CS 598 CM), **Selected to the "List of Teachers Ranked as Excellent By Their Students".**

Fall 2021 - onwards Instructor for Advanced Compiler Technologies Seminar (CS 591 ACT): organized this seminar every semester except Spring 2023, Fall 2024.

External Service

Study Lead **DARPA ISAT Study Co-Lead**

- Co-led the study on ML Optimized Compilers for Heterogeneous Architectures (MOCHA) - Fall 2021 - Summer 2022
- Was instrumental in the inception of the DARPA MOCHA program in 2024

Program Chair **LCPC Program Co-Chair**

Co-chaired the 35th International Workshop on Languages and Compilers for Parallel Computing (LCPC) - October 2022

Debates **ASPLOS 2024**

Part of the ASPLOS debate on, "Should everyone work on AI?" featuring prominent senior researchers.

Mentoring **SIGPLAN-M**

Mentored two female researchers in 2022 by participating in the SIGPLAN-M mentoring program. One went on to become a faculty member at a reputed university.

Program Committees *At Illinois, I have served in 24 program committees in top-tier conferences in programming languages, compilers, architecture, and systems. I have also served on 9 program committees in top-tier machine learning venues.*

- 2027
  - ASPLOS Program Committee
  - EuroSys Program Committee
- 2026
  - PPoPP Program Committee
  - MICRO Program Committee
  - OOPSLA Review Committee
  - HPCA Program Committee
  - ISCA Program Committee
  - EuroSys Program Committee
  - MLSys Program Committee
  - ICML Program Committee
- 2025
  - ASPLOS Program Committee
  - HPCA Program Committee
  - MLSys Program Committee
  - CC Program Committee
  - MICRO Program Committee
  - PACT Program Committee
  - NeurIPS Program Committee
- 2024
  - ASPLOS Program Committee
  - ISCA Program Committee
  - MLSys Program Committee
  - NeurIPS Program Committee

Program Committees (cntd.)
- 2023
  - MLSys Program Committee
  - PLDI Program Committee
  - NeurIPS Program Committee
  - ICLR Program Committee
- 2022
  - PLDI Program Committee
  - CGO Program Committee
  - PACT Program Committee
  - ICML Program Committee
- 2021
  - PACT Program Committee
  - NeurIPS Program Committee
  - ICLR Program Committee - **Outstanding Reviewer**
  - ICML Program Committee

External Review Committee
- ASPLOS 2023 External Review Committee
- ASPLOS 2022 External Review Committee
- ASPLOS 2021 External Review Committee
- PACT 2020 External Review Committee

Student Competition Reviewing
- PACT 2023 Student Research Competition
- SPLASH 2020 Student Research Competition
- ECOOP 2019 Doctoral Symposium

Artifact Evaluation
CGO Artifact Evaluation Committee (2020)

Journal Reviewing
- JPDC: Journal of Parallel and Distributed Computing (2019)
- TACO: Transactions on Architecture and Code Optimization (2026, 2025, 2019)

Subreviewing
ASPLOS 2019, SPAA 2018, CGO 2018, CGO 2017, CGO 2016, PACT 2019

Internal Service

PhD Committee
Preliminary and Final Examinations, *University of Illinois at Urbana-Champaign*
- Serif Yesil (Preliminary: 10/14/2021, Final: 04/06/2022)
- Edward Hutter (Preliminary: 03/28/2023, Final: 12/19/2023)
- Yifan Zhao (Preliminary: 08/23/2024)
- Hyoungwook Nam (Preliminary: 12/03/2024, Final: 07/03/2025)
- Akash Kothari (Preliminary: 12/06/2024)
- Gerasimos Gerogiannis (Preliminary: 02/18/2025)
- Abdul Rafae Noor (Preliminary: 07/08/2025, Final: 04/01/2026)
- Damitha Lenadora (Preliminary: 01/22/2026, chair)
- Stefanos Baziotis (Preliminary: 02/11/2026, chair)
- Ashitabh Misra (Preliminary: 03/09/2026)
- Avaljot Singh (Preliminary: 04/01/2026, co-chair)

Departmental Committees
- 2025-2026: PhD Launch Committee
- 2024-2025: Graduate Awards Committee
- 2023-2024: Graduate Study Committee
- 2022-2023: Academic Appeals Committee, Graduate Awards Committee, Graduate Admissions Committee (Primary Member)
- 2021-2022: Graduate Admissions Committee (Primary Member), Faculty Recruiting Committee (Secondary Member: ML for Systems)

Seminars
- Co-organizer of CS 591 ACT: Advanced Compiler Technologies (Fall 2021 - now)
- Panelist at the departmental NSF CAREER Workshop (May 2025)
- Invited Seminar at UIUC College of Engineering ISUR Program (Oct. 2023)
- Panelist at the Lunch and Learn REU seminar series (Summer 2022)
- Speaker at the UIUC student conference Reflections and Projections (Sept. 2020)

IIDAI
- IBM-Illinois Accelerator Institute, Model Optimization and Edge Deployment Cluster co-lead (with Mudhakar Srivatsa from IBM), 2025

High-school Mentoring
- In Summer 2023, I mentored two female high-school students from Champaign Central High School on the topic of "Optimizing Neural Network Computations". This was done under the K-12 BPC initiative of the CS department.

Invited Talks

Agile and evolvable software construction in the era of rapidly evolving hardware accelerator designs

Cornell CSL Seminar	May. 2026
EMC2 Workshop co-located with ASPLOS	Mar. 2026
Stanford Graduate Seminar	Feb. 2026
Microsoft RiSE seminar	Feb. 2026
UCSC CSE Colloquium	Feb. 2026
Georgia Tech CRNCH Summit	Feb. 2026
CMU ECE Graduate Seminar	Feb. 2026
Google	Dec. 2025
Amazon	Dec. 2025
MIT Fast Code Seminar	Nov. 2025
UIUC Compiler Seminar	Oct. 2025

Making compilers more evolvable with learned cost models

SPIRAL Tutorial co-located with ASPLOS	Mar. 2026
AICoM (AI-Enhanced Co-Design for Next Generation Microelectronics)	Apr. 2024
Amazon AWS	Oct. 2023
AMD Research	Oct. 2023
Intel	Oct. 2023
ACE center theme meeting	Nov. 2023

Towards Automated Construction of Compiler Optimizations

Programming Systems Research Forum (**Keynote Talk**)	Feb. 2022
ML for Computer Architecture and Systems (**Keynote Talk**)	Jun. 2021

Optimizations and abstractions for Sparse Machine Learning

Sparse Workshop co-located with PLDI 2025	Jun. 2025
JUMP 2.0 e-workshop	Jul. 2024

Modernizing Compiler Technology using Machine Learning

Reflections and Projections Conference	Sep. 2020
Workshop on Program Synthesis for Scientific Computing	Aug. 2020
ADA Center	May 2020
Microsoft Research, Redmond	Apr. 2020
University of Southern California, Special Seminar	Apr. 2020
University of Illinois at Urbana-Champaign (UIUC), Special Seminar	Apr. 2020
UC Berkeley, Special Seminar	Mar. 2020
Google Brain, CA	Feb. 2020
MIT Fast Code Seminar	Nov. 2019
Microsoft Research, Redmond	Nov. 2019

Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks
ARM Sep. 2020

goSLP: Globally Optimized Superword Level Parallelism Framework
LLVM Seminar, MIT Aug. 2019
Reservoir Labs, New York Feb. 2019
MIT PL Seminar Mar. 2018

Helium: Lifting High-performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code
University of Moratuwa, Sri Lanka Jan. 2016
MIT Graphics Group Lunch Oct. 2015
Microsoft Research, Redmond Jun. 2015
MIT PL Seminar Jun. 2015

Conference and Other Talks

Modernizing Compiler Technology using Machine Learning
MIT Thesis defense May 2020

Ithemal: Accurate, Portable and Fast Basic Block Throughput Estimation using Deep Neural Networks
International Conference on Machine Learning (ICML) Jun. 2019
ML for Systems Workshop (co-located with ISCA 2019) Jun. 2019

Revec: Program Rejuvenation through Revectorization
International Conference on Compiler Construction (CC) Feb. 2019

goSLP: Globally Optimized Superword Level Parallelism Framework
Object Oriented Programming, Systems, Languages and Applications (OOPSLA) Nov. 2018
LLVM developer meeting. San Jose (poster) Oct. 2018

Helium: Lifting High-performance Stencil Kernels from Stripped x86 Binaries to Halide DSL Code
MIT PL Offsite May 2016
Programming Languages Design and Implementation (PLDI) Jun. 2015