Saswat Padhi

Curriculum Vitae • Last modified on April 5, 2025

spadhi [*at*] google [*dot*] com • Google SJC-TM-2 • San Jose, CA 95126 (USA)
saswatpadhi [*at*] protonmail [*dot*] com • saswat.padhi.me • github.com/SaswatPadhi

Interests

I consider myself a *systems generalist* with a strong interest in low-level software architecture and performance optimizations. I hold a doctoral degree that explores the intersection of data-driven inference techniques with traditional program analysis and verification methods. More recently, my industry work has centered on efficiency and reliability analyses of operating system components.

Employment

Google LLC Sep 2022 – Present
Senior Software Engineer *San Jose, CA*

Android & ChromeOS - Virtualization

- Building compilation infrastructure for guest kernels and OS images, and investigating boot performance for the Linux (Debian) development environment on Android
- Developed user-space guest agents towards replacing legacy LXD-based containers with virtual machines for Linux on ChromeOS

ChromeOS - Performance

- Led the *performance analysis & tiering* project: designed a technique to predict user experience metrics from Chromebook hardware specifications, and deployed it as a internal tool
- Published paper: *Predicting UX on Laptops from Hardware Specifications* (NeurIPS-MLSys 2023)
- Published patent: pending USPTO publication of application
- Mentored 1 PhD intern

Amazon Inc. Aug 2020 – Sep 2022
Applied Scientist II *Boston, MA*

Automated Reasoning Group (ARG)

- Led the *inductive proofs* project: added formal verification support for `C` code with loops and delivered memory safety proofs for AWS projects including FreeRTOS, s2n, and C Commons
- Collaborated with AWS IoT team on static analysis of their events monitoring systems
- Published paper: *Automated Analyses of IoT Event Monitoring Systems* (CAV 2023)
- Published patents: 2024 grant `US 12093160 B1`, applications `US 20240403186 A1`
- Mentored 5 PhD interns; interviewed 30+ applied science and software engineering candidates

Microsoft Corp. (contract via Populus Group) Oct 2017 – Aug 2018
(Part-Time) Research Software Development Engineer *Remote (US)*

Research in Software Engineering (RiSE)

- Developed a technique based on convolutional neural networks to identify data frames in Excel spreadsheets with near-human accuracy

- Prototyped a program synthesis method for automatically converting data cells to formulas
- Published patents: 2024 grant US 12039257 B2, and applications EP 3821366 A1, CN 112424784 A, IN 202117000686 A, and US 20250068837 A1

Education

University of California, Los Angeles Sep 2014 – Jun 2020
Department of Computer Science *Los Angeles, CA*

- Philosophiae Doctor (Ph.D.) · CGPA: 3.8/4.0
- *Specialization:* Programming Languages & Software Systems
- *Dissertation:* Data-Driven Learning of Invariants and Specifications
- *Advisor:* Professor Todd Millstein
- *Outstanding Research in Computer Science* award and *Dissertation-Year Fellowship* from UCLA

Indian Institute of Technology, Bombay Aug 2010 – Apr 2014
Department of Computer Science and Engineering *Mumbai, India*

- Bachelor of Technology (B.Tech.) · CPI: 8.9/10.0
- *Undergraduate Thesis:* Lazy Static Slicing of Functional Programs
- *Advisors:* Professor Amitabha Sanyal, Professor Uday Khedker
- Graduated with *departmental honors* in computer science

D. A. V. Public School, Pokhariput 2004 – 2009
Secondary & Senior Secondary (Science Stream) *Bhubaneswar, India*

- 2009 All India Senior School Certification Examination in Science · Score: 96.0 %
- 2007 All India Secondary School Examination · Score: 98.0 %

Awards

2020	*ACM SIGPLAN Distinguished Paper* award at the 41st PLDI Conference for "*Data-Driven Inference of Representation Invariants*" paper (awarded to 4 out of 77 accepted papers out of 350 total submissions)
2020	*Outstanding Research in Computer Science* award from the CS department at the University of California, Los Angeles (UCLA)
2019 – 2020	*Dissertation-Year Fellowship* (now called Dissertation-Year Award) from UCLA
2018	Gold medal at the 8th *Federated Logic Conference (FLoC)* in London, UK for multiple consecutive victories in the annual SyGuS Competition
2017, 2018	Winner of the annual *Syntax-Guided Synthesis (SyGuS) Competition*
2017 – 2019	*Ph.D. Fellowship* from Microsoft Research (awarded to 10 doctoral candidates across the US and Canada)
2013	Winner of the *Prezi Scale Contest* hosted by HackerRank
2010 – 2014	FIITJEE Scholarship for achieving *All-India Rank 43 in IIT-JEE*
2008	*Kishore Vaigyanik Protsahan Yojana (KVPY) Fellowship* from the Department of Science and Technology, Government of India (awarded to around 100 senior secondary students across India)

2008, 2007 Among the top 25 finalists in the Indian National Astronomy Olympiad (INAO)

2007 *National Talent Search Examination (NTSE) Scholarship*
from the National Council of Education Research and Training (NCERT), India

2007 *Double gold medals* in the International Assessment for Indian Schools (IAIS)
from the University of New South Wales (UNSW), Australia
(for scoring the highest marks in both mathematics and computer science)

2005 Silver medal (All-India Rank 2) in the National Science Olympiad (NSO), India

Patents

Predicting user experience on computing devices from hardware specifications.
Saswat Padhi, Alex Bergman, Allan Knies, Sunil Bhasin, Udaya Kiran Ammu

- Patent rights owned by *Google*
- 2024 Application: US pending publication
- A gradient-boosted regression method for accurately predicting user experience metrics based on the hardware specifications of ChromeOS devices.
- Implemented as an internal tool for evaluating Chromebook laptops.

IoT Event detector correctness verification.
Vaibhav Bhusan Sharma, Andrew Jude Gacek, Michael William Whalen, *Saswat Padhi*, Andrew Apicelli, Raveesh Yadav, Samuel Bayless, Roman Pruzhanskiy, Rajat Gupta, Harshil Rajeshkumar Shah, Fernando Dias Pauer, Ankush Das, Dhivashini Jaganathan

- Patent rights owned by *Amazon*
- 2024 Grant: US 12093160 B1 · 2024 Application: US 20240403186 A1
- A technique for verifying correctness of AWS IoT event detectors that models the event detectors as finite state machines and correctness properties as LTL constraints on them.
- Deployed as an analyzer in AWS IoT Events production service since December 2021.

Systems, methods, and computer-readable media for improved table identification using a neural network.
Benjamin Goth Zorn, Marc Manuel Johannes Brockschmidt, Pallavi Choudhury, Oleksandr Polozov, Rishabh Singh, *Saswat Padhi*

- Patent rights owned by *Microsoft*
- (US Patents) 2024 Grant: US 12039257 B2 · 2025 Application: US 20250068837 A1
- (EU Patents) 2021 Application: EP 3821366 A1
- (Asian Patents) 2021 Application: CN 112424784 A · 2021 Application: IN 202117000686 A
- A neural algorithm, based on convolutional neural networks (CNNs), for identifying regions of tabular data (also called *data frames*) within spreadsheets.
- Implemented as an extension for Microsoft Excel (part of Microsoft Office suite).

Record profiling for dataset sampling.
Daniel G. Simmons, Kevin David James Grealish, Sumit Gulwani, Ranvijay Kumar, Kevin Michael Ellis, *Saswat Padhi*

- Patent rights owned by *Microsoft*
- 2020 Grant: US 10846298 B2

- An algorithm for sampling small, representative subsets of strings from very large datasets based on the *syntactic profiles* of the records in the dataset.
- Implemented in the `Matching.Text` library, publicly available as part of Microsoft PROSE SDK.

Syntactic profiling of alphanumeric strings.
Sumit Gulwani, Prateek Jain, Daniel Adam Perelman, *Saswat Padhi*, Oleksandr Polozov

- Patent rights owned by *Microsoft*
- 2019 Grant: US 10394874 B2 · 2021 Grant: US 11210327 B2
- An algorithm for generating a *syntactic profile* for a collection of strings: a set of *RegExp*-like patterns, each of which describes the structure of a subset of strings.
- Implemented in the `Matching.Text` library, publicly available as part of Microsoft PROSE SDK.

Publications

Journal Articles

1. **FlashProfile: A Framework for Synthesizing Data Profiles**. *OOPSLA 2018*
   Saswat Padhi, Prateek Jain, Daniel Perelman, Alex Polozov, Sumit Gulwani, Todd Millstein
   In the Proceedings of the ACM on Programming Languages (PACMPL) Vol. 2, Object-Oriented Programming, Systems and Applications issue, Article 150, 2018. 28 pages.
   DOI: 10.1145/3276520

Conference Proceedings

1. **Data-Driven Inference of Representation Invariants**. *PLDI 2020*
   Anders Miltner, *Saswat Padhi*, Todd Millstein, David Walker
   (ACM SIGPLAN Distinguished Paper)
   In the proceedings of the 41st ACM SIGPLAN Conference on Programming Language Design and Implementation, London, UK, June 15–20, 2020. 19 pages.
   DOI: 10.1145/3395638 · Extended version: arxiv.org/abs/2003.12106

2. **Overfitting in Synthesis: Theory and Practice**. *CAV 2019*
   Saswat Padhi, Todd Millstein, Aditya Nori, Rahul Sharma
   In the proceedings of the 31st International Conference on Computer Aided Verification, Part I, New York City, NY, USA, July 15-18, 2019. 19 pages.
   DOI: 10.1007/978-3-030-25540-4_17 · Extended version: arxiv.org/abs/1905.07457

3. **A Static Slicing Method for Functional Programs and Its Incremental Version**. *CC 2019*
   Prasanna Kumar K., Amitabha Sanyal, Amey Karkare, *Saswat Padhi*
   In the proceedings of the 28th International Conference on Compiler Construction, Washington, DC, USA, February 16–17, 2019. 12 pages.
   DOI: 10.1145/3302516.3307345

4. **Data-Driven Precondition Inference with Learned Features**. *PLDI 2016*
   Saswat Padhi, Rahul Sharma, Todd Millstein
   In the proceedings of the 37th ACM SIGPLAN Conference on Programming Language Design and Implementation, Santa Barbara, CA, USA, June 13–17, 2016. 19 pages.
   DOI: 10.1145/2908080.2908099

Workshop Papers

1. **Predicting User Experience on Laptops from Hardware Specifications**. *NeurIPS 2023*
   Saswat Padhi, Sunil Bhasin, Udaya Kiran Ammu, Alex Bergman, Allan Knies *(ML4Sys)*
   (Invited for Oral Spotlight Presentation)
   In the proceedings of the 2023 Workshop on Machine Learning for Systems at the Conference on Neural Information Processing Systems, New Orleans, LA, USA, December 16, 2023. 9 pages.
   Available at openreview.net/forum?id=mHShSE7MSU.

2. **OASIS: ILP-Guided Synthesis of Loop Invariants**. *NeurIPS 2020*
   Sahil Bhatia, *Saswat Padhi*, Nagarajan Natarajan, Rahul Sharma, Prateek Jain *(CAP)*
   In the proceedings of the 2020 Workshop on Computer-Assisted Programming at the Conference on Neural Information Processing Systems (Virtual), December 12, 2020. 5 pages.
   Available at openreview.net/forum?id=T591RKxIh6Q.

Industrial Experience Papers

1. **Automated Analyses of IoT Event Monitoring Systems**. *CAV 2023*
   Andrew Apicellii, Sam Bayless, Ankush Das, Andrew Gacek, Dhiva Jaganathan, *Saswat Padhi*, Vaibhav Sharma, Michael W. Whalen, Raveesh Yadav
   In the proceedings of the 35th International Conference on Computer Aided Verification, Part I, Paris, France, July 17-22, 2023. 13 pages.
   DOI: 10.1007/978-3-031-37706-8_2

Theses

1. **Data-Driven Learning of Invariants and Specifications**. *Doctoral Dissertation 2020*
   Saswat Padhi
   Supervised by Prof. Todd Millstein, Department of Computer Science, University of California, Los Angeles, CA. ProQuest ID: Padhi_ucla_0031D_19057. 157 pages.
   DOI: escholarship.org/uc/item/3k89r896

Technical Reports

1. **The SyGuS Language Standard Version 2.1**. *SYNT 2021*
   Saswat Padhi, Elizabeth Polgreen, Mukund Raghothaman, Andrew Reynolds, Abhishek Udupa
   Discussed at the 2021 Workshop on Synthesis at the 33rd International Conference on Computer Aided Verification, Los Angeles, CA, USA, July 19, 2023. 36 pages.
   Available at arxiv.org/abs/2312.06001.

2. **SyGuS-Comp 2018: Result and Analysis**. *SYNT 2018*
   Rajeev Alur, Dana Fisman, *Saswat Padhi*, Rishabh Singh, Abhishek Udupa
   Discussed at the 2018 Workshop on Synthesis at the 8th Federated Logic Conference (FLoC), Oxford, UK, July 18, 2018. 18 pages.
   Available at arxiv.org/abs/1904.07146.

3. **LoopInvGen: A Loop Invariant Generator based on Precondition Inference**. *SYNT 2017 – 19*
   Saswat Padhi, Rahul Sharma, Todd Millstein
   Contribution to the 2017 – 2019 Syntax-Guided Synthesis (SyGuS) competitions. 4 pages.
   Available at arxiv.org/abs/1707.02029.

Visiting Positions

Princeton University Apr 2019 – Jun 2019
Visiting Research Collaborator *Princeton, NJ, USA*

- Collaboration with Prof. David Walker's team in the Programming Languages group
- Extended my prior work on invariant synthesis to recursive data types
- Resulting paper: *Data-Driven Inference of Representation Invariants* (PLDI 2020)

Microsoft Research Lab - India Sep 2018 – Mar 2019
Ph.D. Research Intern *Bengaluru, India*

- Systems Research group, with Dr. Rahul Sharma
- Explored an intersection of my prior work on invariant synthesis with machine learning theory
- Resulting paper: *Overfitting in Synthesis: Theory and Practice* (at CAV 2019)

Microsoft Research Jun 2017 – Oct 2017
Ph.D. Research Intern *Redmond, WA, USA*

- Research in Software Engineering (RiSE) group, with Dr. Ben Zorn and Dr. Rishabh Singh
- Developed a neural (CNN-based) approach for identifying tables (data frames) in Excel sheets
- Resulting patents: 2024 grant US 12039257 B2, and applications EP 3821366 A1, CN 112424784 A, IN 202117000686 A, and US 20250068837 A1

Microsoft Jun 2016 – Dec 2016
Software Engineering Intern *Redmond, WA, USA*

- Program Synthesis using Examples (PROSE) group, with Dr. Sumit Gulwani
- Designed program synthesis techniques for pattern-based profiling of large-scale datasets
- Resulting paper: *FlashProfile: A Framework for Synthesizing Data Profiles* (OOPSLA 2018)
- Resulting granted patents: 2019 US 10394874 B2, 2020 US 10846298 B2 and 2021 US 11210327 B2

Google Inc. May 2013 – Jul 2013
Summer Intern *Mountain View, CA, USA*

- Technical Infrastructure (TI) group, reporting to Smeeta Jalan
- Worked on test automation for Google's Borg and Omega cluster-management systems

Technische Universität Braunschweig (TU-Br) May 2012 – Jul 2012
Summer Research Intern *Braunschweig, Germany*

- Institut für Informationssysteme (IFIS), advised by Prof. Dr. Wolf-Tilo Balke
- Comparative analysis of bibliometric versus semantic measures for estimating topical similarity of scientific publications (primarily the PubMed corpus)

Invited Presentations

ML-Based User Experience Prediction

- Dec 2023: ML4Sys Workshop at 37th NeurIPS Conference, New Orleans, LA, USA.

Data-Driven Precondition Inference

- Sep 2020: Amazon Web Services, Inc. (Virtual)
- Jan 2017: Microsoft Research Lab, Bengaluru, India.

- Jun 2016: 37th ACM SIGPLAN PLDI Conference, Santa Barbara, CA, USA.
- May 2016: University of California at Berkeley, CA, USA.
- May 2016: Software Lunch Talk, Stanford University, CA, USA.
- Dec 2015: 15th SoCal PLS Workshop, Pomona College, CA, USA.

Overfitting in Program Synthesis

- Jul 2019: 31st CAV Conference, New York City, NY, USA.

Synthesizing Pattern-Based Data Profiles

- Nov 2018: 18th ACM SIGPLAN SPLASH-OOPSLA Conference, Boston, MA, USA.
- Oct 2018: Microsoft Research Lab, Bengaluru, India.
- Dec 2016: Microsoft Research, Redmond, WA, USA.

Professional Service

Organizing Committee Member

- 2019 – 2021 Syntax-Guided Synthesis Competition (SyGuS-Comp)

Program / Review Committee Member

- 2024 Workshop on Horn Clauses for Verification and Synthesis (HCVS)
  at the European Joint Conferences on Theory and Practice of Software (ETAPS)
- 2022 Workshop on Horn Clauses for Verification and Synthesis (HCVS)
  at the European Joint Conferences on Theory and Practice of Software (ETAPS)
- 2021 Conference on Programming Languages Design and Implementation (PLDI)
- 2020 Conference on Programming Languages Design and Implementation (PLDI)
- 2021 Workshop on Synthesis (SYNT)
  at the International Conference on Computer-Aided Verification (CAV)
- 2019 Workshop on Debugging Machine Learning Models (DebugML)
  at the International Conference on Learning Representations (ICLR)

Artifact Evaluation Committee Member

- 2020 Conference on Principles of Programming Languages (POPL)
- 2019 Static Analysis Symposium (SAS)
- 2019 Object-Oriented Programming, Systems, and Applications (OOPSLA) track
  of the Conference on Systems, Programming, Languages, and Applications:
  Software for Humanity (SPLASH)
- 2018 Object-Oriented Programming, Systems, and Applications (OOPSLA) track
  of the Conference on Systems, Programming, Languages, and Applications:
  Software for Humanity (SPLASH)

Invited External Reviewer

- 2024 Journal of Artificial Intelligence Research (JAIR)
- 2022 International Conference on Foundations of Software Science and Computation Structures

- 2021 IEEE Transactions on Software Engineering (TSE)
- 2019 International Conference on Computer-Aided Verification (CAV)
- 2019 ACM SIGSOFT India's Innovations in Software Engineering Conference (ISEC)

Mentored Students

- Tomoya Matsubara, intern at Google during doctoral study at Keio University, Japan.
  (Summer 2024) Investigating performance impact of ChromeOS kernel security flags
- Dr. Akshay Utture, intern at AWS during doctoral study at UCLA.
  (Summer 2022) Proving memory-safety of the task queue in the FreeRTOS kernel
- Dr. Imtiaz Kareem, intern at AWS during doctoral study at Purdue University.
  (Summer 2021, co-mentored with Dr. Vaibhav Sharma) Type checking for AWS-internal DSLs
- Long Pham, intern at AWS during doctoral study at CMU.
  (Summer 2021) Adding termination proofs to the C Bounded Model Checker (CBMC)
- Dr. Aalok Thakkar, intern at AWS during doctoral study at UPenn.
  (Summer 2021) Adding history variables to the C Bounded Model Checker (CBMC)
- Dr. Pamina Georgiou, intern at AWS during doctoral study at TU-Wein, Austria.
  (Fall 2020) Proving correctness of AWS projects using the C Bounded Model Checker (CBMC)
- Adam Stein, during undergraduate study at UCLA.
  (2019 – 2020) Adding Bit Vector theory to the LoopInvGen synthesizer
- Sahil Bhatia, Research Fellow at Microsoft.
  (2018 – 2019) Synthesizing loop invariants using integer linear programming (ILP)
- Brett Chalabian, during graduate studiy at UCLA.
  (2017 – 2019) Adding user-specified ad-hoc features in the LoopInvGen synthesizer
- Zhouheng (Jeffrey) Sun, during undergraduate study at UCLA.
  (2017 – 2019) Adding higher-order features to the LoopInvGen synthesizer

Teaching Experience

Programming Languages CS 131
Graduate Teaching Assistant *University of California, Los Angeles*

- Spring 2016 · Instructor: Prof. Todd Millstein
- Fall 2014 · Instructor: Prof. Todd Millstein

Abstractions and Paradigms in Programming CS 152
Undergraduate Teaching Assistant *Indian Institute of Technology, Bombay*

- Spring 2014 · Instructor: Prof. Amitabha Sanyal

Computer Programming and Utilization Fall 2011 · Fall 2013
Undergraduate Teaching Assistant *Indian Institute of Technology, Bombay*

- Fall 2013 · Instructor: Padma Shri Prof. Deepak B. Phatak
- Fall 2011 · Instructor: Padma Shri Prof. Deepak B. Phatak

Software Releases

pseudocode.js 2020 – Present
[github.com/SaswatPadhi/pseudocode.js] *MIT License*

- Current lead developer; originally authored by Tate Tian.
- A JavaScript library that accurately typesets LaTeX-style pseudocodes to HTML.

LoopInvGen 2016 – 2020
[github.com/SaswatPadhi/LoopInvGen] *MIT License*

- Lead developer; co-developed with Todd Millstein and Rahul Sharma.
- Infers necessary preconditions and sufficient loop invariants, fully automatically.
- Won SyGuS-Comp 2017, 2018. Second place in SyGuS-Comp 2019.

Personal Information

- Citizenship: Indian
- Residence: United State of America
  approved EB-1A permanent residence petition, pending adjustment of status
- Languages: Odiya (native), English (fluent), Hindi (fluent), Sanskrit (basic)