Aditya G. Parameswaran

2165 Bakar Gateway, 2300 Hearst Ave
Berkeley CA 94720
University of California, Berkeley

Email: adityagp @ berkeley.edu
Website: adityagp.net

Employment

- **University of California, Berkeley**, Berkeley, CA.
  Associate Professor, Electrical Engineering and Computer Sciences. 2021—
  Assistant Professor, Electrical Engineering and Computer Sciences. 2019—2021
  Assistant/Associate Professor, School of Information. 2019—2021; 2021—2024
  Co-Director, California Police Records Project. 2023—
  Co-Director, EPIC Data Lab, 2021—
  Faculty Affiliate, Berkeley Institute of Data Science (BIDS). 2022—
  Faculty Affiliate, Berkeley Institute of Design (BID). 2019—
- **Snowflake, Inc.**, San Mateo, CA.
  AI x Data Research Lead (Part-time). 2026—
  Leading research efforts on AI and data systems.
- **Bauplan, Inc.**, New York, NY.
  Technical and Strategic Advisor (Advisory, Part-time). 2025—
- **Ponder, Inc. (Acq., Snowflake)**, Berkeley, CA.
  Co-founder and President. 2021—2023
  Ponder commercialized open-source data science tools developed as part of research (Modin, Lux). Raised 7M in 2021, led by top-tier VCs (Lightspeed, 8VC, Intel), and grew to 15 employees. Acquired by Snowflake to accelerate their data science offerings on October 24, 2023. Articles from Venture Beat and Snowflake.
- **University of Illinois, Urbana-Champaign**, Urbana, IL.
  Assistant Professor, Computer Science. 2014—2019
  Affiliate, Informatics Institute, Institute for Genomic Bio. (IGB), Beckman Institute. 2016—2019
- **Massachussetts Institute of Technology**, Cambridge, MA.
  Postdoctoral Researcher, Computer Science and Artifical Intelligence Lab. 2013—2014
- **Microsoft Research New England**, Cambridge, MA.
  Consulting Researcher. 2013—2014

Education

- **Stanford University**, Stanford, CA.
  Ph.D. in Computer Science. 2007—2013
  Advisor: Hector Garcia-Molina; Title: *Human-Powered Data Management*
  Winner: **ACM SIGMOD Jim Gray Dissertation Award**, **ACM SIGKDD Dissertation Award Runner Up**, **Stanford's Arthur Samuels Dissertation Award**
- **IIT Bombay**, Mumbai, India.
  B.Tech in Computer Science. 2003—2007

Selected Honors and Awards

- **TCDE Ten-Year Influential Paper Award** for "Interactive Data Exploration with Smart Drill-Down", awarded to a single paper from the ICDE 2016 conference with the most impact, 2026
- **CHI 2026 Best Paper Award** for "RAG Without the Lag", awarded to $<1\%$ of submissions, 2026
- **UIST 2025 Best Paper Honorable Mention** for "DocWrangler", awarded to $<1\%$ of submissions, 2025
- Named a **Kavli Fellow** by the National Academy of Sciences, 2023
- **Young Alumni Achiever Award** from the Indian Institute of Technology (IIT), Bombay, 2022
- **SIGMOD 2020 Best Paper Award** for my paper "ShapeSearch: A Flexible and Efficient System for Shape-based Exploration of Trendlines", awarded to two papers out of 144 accepted research papers and 450 submissions, 2020

- ◇ **Alfred P. Sloan Fellowship** in Computer Science, awarded to 20 early career researchers each year "in recognition of distinguished performance and a unique potential to make substantial contributions to their field", 2020
- ◇ **VLDB Early Career Research Contributions Award**, from the Very Large Data Bases (VLDB) Endowment, recognizing demonstrated research impact through a specific technical contribution of high significance, for up to eight years post Ph.D. *"for developing tools for large-scale data exploration, targeting non-programmers"*, 2019
- ◇ **Best Demo Award**, ICDE for my paper "Faster, Higher, Stronger: Redesigning Spreadsheets for Scale". Awarded to two out of 24 demonstrations, out of many more submissions, 2019
- ◇ Army Research Office **Young Investigator Program (ARO YIP) Award**, 2018
- ◇ IEEE **TCDE Early Career Award** (now called Rising Star Award), from the Technical Committee on Data Engineering, for contributions to the field of databases, for up to five years post Ph.D. *"for developing new interactive tools & techniques that expand the reach of data analytics"*, 2017
- ◇ **NSF CAREER** Award, 2017
- ◇ Selections for **best-papers-of-conference** for the following conferences (awarded to ∼5 papers per year, out of hundreds of accepted papers):
  - · **SIGMOD'26**, for my paper "Cut Costs, Not Accuracy: LLM-Powered Data Processing with Guarantees"
  - · **VLDB'23**, for my paper "Transactional Panorama: A Conceptual Framework for User Perception in Analytical Interfaces"
  - · **VLDB'17**, for my paper "OrpheusDB: Bolt-on Versioning for Relational Databases"
  - · **AISTATS'17**, for my paper "On the Interpretability of Conditional Probability Estimates"
  - · **ICDE'16**, for my paper "Interactive Data Exploration with Smart Drill-Down"
  - · **ICDE'14**, for my paper "Crowd-Powered Find Algorithms"
  - · **KDD'12**, for my paper "Active Sampling for Entity Matching"
  - · **VLDB'10**, for my paper "Towards the Web of Concepts: Extracting Concepts from Large Datasets"

Other Honors and Awards

- ◇ **SIGMOD 2026 Distinguished Associate Editor**, 2026
- ◇ **CACM Research Highlight Award**, 2022
- ◇ **CIDR Gong Show Award Honorable Mention**, 2022
- ◇ Runner Up, **Facebook Probability and Programming Research Award**, 2019
- ◇ Dean's Award for **Excellence in Research**, from the College of Engineering at the University of Illinois, 2018
- ◇ C.W. Gear **Outstanding Junior Faculty Award**, from the Computer Science dept. at the University of Illinois, 2017
- ◇ Listed in **"Instructors rated as Excellent by students @ Illinois"**, 2017, 2015
- ◇ Google **Faculty Research Award**, 2015, and **Focused Research Award**, 2017
- ◇ **NeurIPS'26** Oral for my paper "Inductive Deductive Synthesis: Enabling AI to Generate Formally Verified Systems"
- ◇ **NeurIPS'25** Spotlight for my paper "Why Do Multi-Agent LLM Systems Fail?"
- ◇ Best Demo Paper Honorable Mention at **SIGMOD'17** for "OrpheusDB: A Light-weight Approach to Relational Dataset Versioning".
- ◇ Selected as a **ACM Heidelberg Laureate**, 2013
- ◇ **Yahoo! Key Scientific Challenges Award, 2010**, for my proposal, "Information-Powered Recommendations"

Relevant Expertise

My work appears in top venues in data management, human-computer interaction, and artificial intelligence/natural language processing. I've published over **100** papers in top-tier venues with an **h-index of 57**, an **i-10 index of 120**, and **10,000+** citations[1].

Publications

◇ **Refereed Full-length Papers** (Students I advise or co-advise are marked using a ⋆.)

126. S. Agarwal⋆, A. Biswal, S. Zeighami⋆, A. Cheung, J. Gonzalez, **A. Parameswaran**. Arming Data Agents with Tribal Knowledge. *VLDB'27: 53rd Int'l Conf. on Very Large Data Bases*, Athens, Greece, August 2027.

[1] Any metric, especially those based on paper or citation counting, is flawed; we present these statistics as is without judgment.

125. Y. Sun*, S. Zeighami*, B. Chopra*, S. Shankar*, **A. Parameswaran**. Semantic Data Processing with Holistic Data Understanding. *VLDB'27: 53rd Int'l Conf. on Very Large Data Bases*, Athens, Greece, August 2027.
124. R. Ma*, Y. Lin*, **A. Parameswaran**. Robust Hierarchical Structures for Agentic Document Analysis. *SIGMOD'27: ACM SIGMOD Int'l Conf. on Management of Data*, Huntington Beach, USA, June 2027.
123. Q. Mang, Y. Xiang, H. Zhou, R. He, J. Yu, H. Li, **A. Parameswaran**, A. Cheung. PLOP: Cost-Based Placement of Semantic Operators in Hybrid Query Plans. *SIGMOD'27: ACM SIGMOD Int'l Conf. on Management of Data*, Huntington Beach, USA, June 2027.
122. S. Agarwal*, A. Krentsel, S. Liu, M. Cemri, A. Cheng, R. Meng, T. Pfister, C. Li, S. Ratnasamy, **A. Parameswaran**, M. Zaharia, I. Stoica, M. Lesani. Inductive Deductive Synthesis: Enabling AI to Generate Formally Verified Systems. *NeurIPS'26: 40th Advances in Neural Information Processing Systems*, Sydney, Australia, December 2026. **Oral Presentation.**
121. B. Chopra*, M. Chen, R. Dang, C. Park, S. Shankar*, S. Zeighami*, B. Hartmann, **A. Parameswaran**. Who's Keeping Score? Interactive Steering of LLM-Powered Scoring with Attune. *UIST'26: 39th ACM Symposium on User Interface Software and Technology*, Detroit, USA, November 2026.
120. R. Ma*, S. Shankar*, R. Chen*, Y. Lin*, S. Zeighami*, R. Ghosh, A. Gupta, A. Gupta, T. Gopal, **A. Parameswaran**. Can AI Agents Answer Your Data Questions? A Benchmark for Data Agents. *EMNLP'26: 31st Conf. on Empirical Methods in Natural Language Processing*, Budapest, Hungary, October 2026.
119. S. Zeighami*, S. Shankar*, **A. Parameswaran**. Featurized-Decomposition Join: Low-Cost Semantic Joins with Guarantees. *VLDB'26: 52nd Int'l Conf. on Very Large Data Bases*, Boston, USA, September 2026.
118. Y. Lin*, S. Zeighami*, **A. Parameswaran**. Bolt-on, Verifiable Provenance for LLM-Powered Data Processing. *VLDB'26: 52nd Int'l Conf. on Very Large Data Bases*, Boston, USA, September 2026.
117. L. Wei*, S. Shankar*, S. Zeighami*, Y. Chung, F. Ozcan, **A. Parameswaran**. Multi-Objective Agentic Rewrites for Unstructured Data Processing. *VLDB'26: 52nd Int'l Conf. on Very Large Data Bases*, Boston, USA, September 2026.
116. Y. Lin*, M. Hasan*, R. Kosalge*, A. Cheung, **A. Parameswaran**. Visual Template Inference for Data Extraction from Documents. *SIGMOD'26: ACM SIGMOD Int'l Conf. on Management of Data*, Bengaluru, India, June 2026.
115. S. Zeighami*, S. Shankar*, **A. Parameswaran**. Cut Costs, Not Accuracy: LLM-Powered Data Processing with Guarantees. *SIGMOD'26: ACM SIGMOD Int'l Conf. on Management of Data*, Bengaluru, India, June 2026. Invited to **Special Issue of TODS for SIGMOD 2026 Best Papers**.
114. S. Shankar*, S. Zeighami*, **A. Parameswaran**. Task Cascades for Efficient Unstructured Data Processing. *SIGMOD'26: ACM SIGMOD Int'l Conf. on Management of Data*, Bengaluru, India, June 2026.
113. S. Shankar*, **A. Parameswaran**. What Do Users Actually Do with LLM-Powered Data Systems? *IEEE Data Engineering Bulletin*, June 2026.
112. Q. Romero Lauro*, S. Shankar*, S. Zeighami*, **A. Parameswaran**. RAG Without the Lag: Enabling 'What-If' Analysis for Retrieval-Augmented Generation Pipelines. *CHI'26: 44th ACM Conf. on Human Factors in Computing Systems*, Barcelona, Spain, April 2026. **Best Paper Award.**
111. S. Liu, S. Ponnapalli, S. Shankar*, S. Zeighami*, A. Zhu, S. Agarwal*, R. Chen*, S. Suwito, S. Yuan, I. Stoica, M. Zaharia, A. Cheung, N. Crooks, J. Gonzalez, **A. Parameswaran**. Supporting Our AI Overlords: Redesigning Data Systems to be Agent-First. *CIDR'26: Conf. on Innovative Data Systems Research*, Chaminade, USA, January 2026.
110. M. Cemri, M. Z. Pan, S. Yang, L. A. Agrawal, B. Chopra*, R. Tiwari, K. Keutzer, **A. Parameswaran**, D. Klein, K. Ramchandran, M. Zaharia, J. E. Gonzalez, I. Stoica. Why Do Multi-Agent LLM Systems Fail? *NeurIPS'25: 38th Advances in Neural Information Processing Systems, Datasets and Benchmarks Track*, Vancouver, Canada, December 2025. **Spotlight paper.**
109. S. Shankar*, B. Chopra*, M. Hasan*, S. Lee, B. Hartmann, J. M. Hellerstein, **A. Parameswaran**, E. Wu. Steering Semantic Data Processing with DocWrangler. *UIST'25: 40th ACM Symposium on User Interface Software and Technology*, Busan, Korea, October 2025. **Best Paper Honorable Mention.**
108. R. Lin*, B. Chopra*, W. Lin*, S. Shankar*, M. Hulsebos, **A. Parameswaran**. Rethinking Dataset Discovery with DataScout. *UIST'25: 40th ACM Symposium on User Interface Software and Technology*, Busan, Korea, October 2025.
107. S. Shankar*, T. Chambers, T. Shah, **A. Parameswaran**, E. Wu. DocETL: Agentic Query Rewriting and

Evaluation for Complex Document Processing. *VLDB'25: 51st Int'l Conf. on Very Large Data Bases*, London, UK, September 2025.

106. Y. Chen, X. Li, J. Tao, L. Ramjit, S. Mitra, J. Ghaderi, R. Netravali, **A. Parameswaran**, D. Rubenstein, E. Wu. Physical Visualization Design: Decoupling Interface and System Design. *SIGMOD'25: ACM SIGMOD Int'l Conf. on Management of Data*, Berlin, Germany, June 2025.
105. D. Tang*, A. Fekete, I. Gupta, **A. Parameswaran**. Transactional Panorama: A Conceptual Framework for User Perception in Analytical Visual Interfaces (Extended Version). *VLDB Journal*, Volume 34, May 2025. Extended Version of VLDB'23 paper as part of the **Special Issue of VLDB Journal for VLDB 2023 Best Papers**.
104. R. Warren*, **A. Parameswaran**, L. Pickoff-White, N. Salehi. RequestAtlas: Supporting the Slow and Iterative Process of Requesting Public Records. *CSCW'25: Conf. on Computer Supported Cooperative Work*, Bergen, Norway, 2025.
103. Y. Lin*, M. Hulsebos*, R. Ma*, S. Shankar*, S. Zeighami*, **A. Parameswaran**, E. Wu. Towards Accurate and Efficient Document Analytics with Large Language Models. *ICDE'25: 41st IEEE Int'l Conf on Data Engineering*, Hong Kong, May 2025.
102. R. Vir*, S. Shankar*, H. Chase, W. Hinthorn, **A. Parameswaran**. PromptEvals: A Dataset of Assertions and Guardrails for Custom Production Large Language Model Pipelines. *NAACL'25: 20th Conf. of the North American Chapter of the Association for Computational Linguistics*, Albuquerque, USA, April 2025.
101. S. Zeighami*, Z. Wellmer*, **A. Parameswaran**. NUDGE: Lightweight Non-Parametric Fine-Tuning of Embeddings for Retrieval. *ICLR'25: 13th Int'l Conf on Learning Representations*, Singapore, March 2025.
100. S. Zeighami*, Y. Lin*, S. Shankar*, **A. Parameswaran**. LLM-Powered Proactive Data Systems. *IEEE Data Engineering Bulletin*, Issue on LLMs-meets-data, March 2025.
99. R. Garcia, P. Kallanagoudar, C. Anand, S. E. Chasins, J. Hellerstein, **A. G. Parameswaran**: Flow with FlorDB: Incremental Context Maintenance for the Machine Learning Lifecycle. *CIDR'25: Conf. on Innovative Data Management*, Amsterdam, 2025.
98. S. Shankar*, R. Garcia*, J. M. Hellerstein, **A. Parameswaran**. "We have no idea how models will behave in production until production": How engineers operationalize machine learning. *CSCW'24: Conf. on Computer Supported Cooperative Work*, Costa Rica, 2024.
97. H. Li*, N. Chalapathi*, A. Cheung, **A. Parameswaran**. Inferring Visualization Intent from Conversation *CIKM'24: Int'l Conf on Information and Knowledge Management*, Boise, USA, 2024.
96. S. Shankar*, J. D. Zamfirescu-Pereira, B. Hartmannn, **A. Parameswaran**, I. Arawjo. Who Validates the Validators? Aligning LLM-Assisted Evaluation of LLM Outputs with Human Preferences. *UIST'24: Int'l Conf on User Interace Software and Technology*, Pittsburgh, USA, 2024.
95. S. Shankar*, H. Li*, P. Asawa*, M. Hulsebos*, Y. Lin*, J. D. Pereira, H. Chase, W. Fu-Hinthorn, **A. Parameswaran**, E. Wu. SPADE: Synthesizing Assertions for Large Language Model Pipelines. *VLDB'24: 50th Intl. Conf. on Very Large Data Bases*, Guangzhou, China, 2024.
94. J. Wu*, D. Tang*, N. Chalapathi*, T. Chambers, J. Ciccolini, C. Philips, L. Pickoff-White, **A. Parameswaran**. Dealing with Acronyms, Abbreviations, and Typos in Real-World Entity Matching. *VLDB'24: 50th Intl. Conf. on Very Large Data Bases*, Guangzhou, China, 2024.
93. M. Hulsebos*, W. Lin*, S. Shankar*, **Aditya G. Parameswaran**. "It Took Longer than I was Expecting": Why is Dataset Search Still so Hard?. *HILDA'24: Human-in-the-loop Data Analytics Workshop* at SIGMOD, Santiago, Chile, 2024.
92. **A. Parameswaran**, S. Shankar*, P. Asawa*, N. Jain*, Y. Wang*. Revisiting Prompt Engineering via Declarative Crowdsourcing. *CIDR '24: Conf. on Innovative Data Management*, Chaminade, USA, 2024.
91. S. Shankar*, L. Fawaz, K. Gyllstrom, **A. Parameswaran**. Automatic and Precise Data Validation for Machine Learning. *CIKM'23: 32nd Int'l Conf on Information and Knowledge Management*, Birmingham, UK, 2023.
90. S. Shankar*, **A. Parameswaran**. Towards Observability for Production Machine Learning Pipelines. *VLDB'23: 49th International Conference on Very Large Data Bases*, Vancouver, Canada, 2023.
89. S. Shankar*, S. Macke, S. Chasins, A. Head, **A. Parameswaran**. Bolt-on, Compact, and Rapid Program Slicing for Notebooks. *VLDB'23: 49th International Conference on Very Large Data Bases*, Vancouver, Canada, 2023.
88. D. Tang*, A. Fekete, I. Gupta, **A. Parameswaran**. Transactional Panorama: A Conceptual Framework for

User Perception in Analytical Visual Interfaces. *VLDB'23: 49th International Conference on Very Large Data Bases*, Vancouver, Canada, 2023. Invited to **Special Issue of VLDB Journal for VLDB 2023 Best Papers**.

87. D. Tang*, F. Chen, C. De Leon, T. Wattanawaroon, J. Yun, S. Seshadri, **A. Parameswaran**. Efficient and Compact Spreadsheet Formula Graphs.. *ICDE'23: 37th Int'l Conf. on Data Engineering*, Anaheim, USA, 2023.
86. D. Lee*, D. Tang*, K. Agarwal*, T. Boonmark*, C. Chen*, J. Kang*, U. Mukhopadhyay*, J. Song*, M. Yong, M. Hearst, **A. Parameswaran**. Lux: Always-on Visualization Recommendations for Exploratory Data Science. *VLDB'22: 48th International Conference on Very Large Data Bases*, Sydney, Australia and Zoom, 2022.
85. D. Petersohn*, D. Tang*, R. Durrani*, A. Melik-Adamyan, J. Gonzalez, A. Joseph, **A. Parameswaran**. Flexible Rule-Based Decomposition and Metadata Independence in Modin: A Parallel Dataframe System. *VLDB'22: 48th International Conference on Very Large Data Bases*, Sydney, Australia and Zoom, 2022.
84. T. Siddiqui, P. Luh, Z. Wang, K. Karahalios, **A. Parameswaran**. Expressive Querying for Accelerating Visual Analytics. *Communications of the ACM*, 2022. Invited paper accompanying the **SIGMOD Best Paper Award**.
83. W. Epperson, D. Lee*, L. Wang, K. Agarwal*, **A. Parameswaran**, D. Moritz, A. Perer. Leveraging Analysis History for Improved In Situ Visualization Recommendation. *EuroVis'22: 24st Eurographics Conference on Visualization*, Rome, Italy, 2022.
82. J. Hellerstein, **A. Parameswaran**. Piloting Data Engineering at Berkeley. Data Ed Workshop at *SIGMOD'22: SIGMOD International Conference on Data Management*, USA, 2022.
81. D. Lee*, V. Setlur, M. Tory, K. Karahalios, **A. Parameswaran**. Deconstructing Categorization in Visualization Recommendation: A Taxonomy and Comparative Study. *InfoVis'21: IEEE Int'l Conf. on Information Visualization*, Zoom, 2021.
80. S. Macke*, H. Gong*, D. Lee*, A. Head, D. Xin*, **A. Parameswaran**. Fine-Grained Lineage for Safer Notebook Interactions. *VLDB'21: 47th Int'l Conf. on Very Large Data Bases*, Copenhagen, Denmark, 2021.
79. S. Rahman*, M. Bendre*, Y. Liu*, S. Zhu*, N. Su*, K. Karahalios, **A. Parameswaran**. NOAH: Interactive Spreadsheet Exploration with Dynamic Hierarchical Overviews. *VLDB'21: 47th Int'l Conf. on Very Large Data Bases*, Copenhagen, Denmark, 2021.
78. D. Xin*, H. Miao, **A. Parameswaran**, N. Polyzotis. Production Machine Learning Pipelines: Empirical Analysis and Optimization Opportunities *SIGMOD Int'l Conf. on Management of Data*, Xi'an, China, 2021
77. D. Xin*, D. Petersohn*, D. Tang*, Y. Wu*, J. Gonzalez, J. Hellerstein, A. Joseph, **A. Parameswaran**. Enhancing the Interactivity of Dataframe Queries by Leveraging Think Time. *IEEE Data Engineering Bulletin*, Issue on Data Validation for Machine Learning, 2021
76. D. Xin*, E. Wu*, D. Lee*, N. Salehi, **A. Parameswaran**. Whither AutoML? Understanding the Role of Automation in Machine Learning Workflows. *CHI'21: Int'l Conf. on Human Factors in Computing Systems*, Yokohama, Japan, 2021.
75. T. Siddiqui, P. Luh, Z. Wang, K. Karahalios, **A. Parameswaran**. From Sketching to Natural Language: Expressive Visual Querying for Accelerating Insight. *SIGMOD Record*, 50(1): 51-58, 2021 Invited paper accompanying the **SIGMOD Best Paper Award**.
74. S. Macke*, M. Aliakbarpour, I. Diakonikolas, **A. Parameswaran**, R. Rubinfeld. Rapid Approximate Aggregation with Distribution-Sensitive Interval Guarantees. *ICDE'21: 37th Int'l Conf. on Data Engineering*, Chania, Greece, 2021
73. D. Lee*, T. Siddiqui*, K. Karahalios, **A. Parameswaran**. Three Lessons from Accelerating Scientific Insight Discovery via Visual Querying. *Patterns*, Cell Press, Volume 1, Issue 7, 100126, 2020.
72. S. Huang*, C. Blatti, S. Sinha, **A. Parameswaran**. Uncovering Effective Explanations for Interactive Genomic Data Analysis. *Patterns*, Cell Press, Volume 1, Issue 6, 100093, 2020.
71. D. Petersohn*, S. Macke*, D. Xin*, W. Ma*, D. Lee*, S. Mo, J. Gonzalez, J. Hellerstein, A. Joseph, **A. Parameswaran**. Towards Scalable Dataframe Systems. *VLDB'20: 46th Int'l Conf. on Very Large Data Bases*, Tokyo, Japan, 2020.
70. S. Rahman*, K. Mack*, M. Bendre*, R. Zhang*, K. Karahalios, **A. Parameswaran**. Benchmarking Spreadsheet Systems. *SIGMOD'20: SIGMOD Int'l Conf. on Management of Data*, Portland, USA, 2020. Covered in the **Morning Paper**, a popular industry blog.
69. T. Siddiqui*, P. Luh*, Z. Wang*, K.Karahalios, **A. Parameswaran**. ShapeSearch: A Flexible and Efficient

System for Shape-based Exploration of Trendlines. *SIGMOD'20: SIGMOD Int'l Conf. on Management of Data*, Portland, USA, 2020. **Best Paper Award (2 out of 450 submissions)**.

68. A. Lee[⋆], D. Xin[⋆], D. Jung-Lin Lee[⋆], **A. Parameswaran**. Demystifying a Dark Art: Understanding Real-World Machine Learning Model Development. *HILDA '20: Workshop on Human-in-the-Loop Data Analytics* at the ACM SIGMOD Int'l Conf. on Management of Data, Portland, USA, 2020.
67. P. Yang[⋆], T. Cheng[⋆], S. Rahman[⋆], M. Bendre[⋆], K. Karahalios, **A. Parameswaran**. Understanding Data Analysis Workflows on Spreadsheets: Roadblocks and Opportunities. *HILDA '20: Workshop on Human-in-the-Loop Data Analytics* at the ACM SIGMOD Int'l Conf. on Management of Data, Portland, USA, 2020.
66. S. Huang[⋆], L. Xu[⋆], J. Liu[⋆], A. Elmore, **A. Parameswaran**. OrpheusDB: Bolt-on Versioning for Relational Databases. *VLDB Journal*, Volume 29 (509-538), January 2020. Extended Version of VLDB'17 paper.
65. T. Rekatsinas, A. Deshpande, **A. Parameswaran**. CRUX: Adaptive Querying for Efficient Crowdsourced Data Extraction. *CIKM '19: 28th Int'l Conf. on Information and Knowledge Management*, Beijing, China, 2019.
64. **A. Parameswaran**. Enabling Data Science for the Majority. *VLDB'19: 45th Int'l Conf. on Very Large Data Bases*, Los Angeles, USA, 2019. Invited paper accompanying the **VLDB Early Career Research Contributions Award**.
63. D. Lee[⋆], J. Lee[⋆], T. Siddiqui[⋆], J. Kim[⋆], K. Karahalios, **A. Parameswaran**. You can't always sketch what you want: Understanding Sensemaking in Visual Query Systems. *VAST'19 at VIS:* IEEE Int'l Conf. on Visual Analytics Science & Technology, Vancouver, Canada, 2019.
62. D. Xin[⋆], S. Macke[⋆], L. Ma[⋆], J. Liu[⋆], S. Song[⋆], **A. Parameswaran**. Helix: Holistic Optimization for Accelerating Iterative Machine Learning. *VLDB'19: 45th Int'l Conf. on Very Large Data Bases*, Los Angeles, USA, 2019. Mentioned as a **Project to Know** in an industry blog post by Amplify Partners.
61. M. Bendre[⋆], T. Wattanawaroon[⋆], K. Mack[⋆], K. Chang, **A. Parameswaran**. Anti-Freeze for Large and Complex Spreadsheets: Asynchronous Formula Computation. *SIGMOD'19: SIGMOD Int'l Conf. on Management of Data*, Amsterdam, The Netherlands, 2019.
60. C. Yen, **A. Parameswaran**, W. Fu. An Exploratory User Study of Visual Causality Analysis. *EuroVis'19: 21st Eurographics Conference on Visualization*, Porto, Portugal, 2019.
59. D. Lee[⋆], H. Dev[⋆], H. Hu[⋆], H. Elmeleegy, **A. Parameswaran**. Avoiding Drill-down Fallacies with VisPilot: Assisted Exploration of Data Subsets. *IUI'19: Int'l Conf. on Intelligent User Interfaces*, Los Angeles, USA, 2019.
58. D. Lee[⋆], S. Macke[⋆], D. Xin[⋆], A. Lee[⋆], S. Huang[⋆], and **A. Parameswaran**. A Human-in-the-loop Perspective on AutoML: Milestones and the Road Ahead, *IEEE Data Engineering Bulletin*, Special Issue of DB4AI and AI4DB, June 2019.
57. M. Joglekar[⋆], H. Garcia-Molina, and **A. Parameswaran**. Interactive Data Exploration with Smart Drill-down (Extended Version). *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, January 2019. Extended Version of ICDE'16 paper.
56. S. Macke[⋆], Y. Zhang[⋆], S. Huang[⋆], **A. Parameswaran**. Adaptive Sampling for Rapidly Matching Histograms. *VLDB'18: 44th Int'l Conf. on Very Large Data Bases*, Rio De Janeiro, Brazil, 2018.
55. Y. Gao[⋆], S. Huang[⋆], **A. Parameswaran**. Navigating the Data Lake with Datamaran: Automatically Extracting Structure from Log Datasets. *SIGMOD '18: ACM SIGMOD Int'l Conf. on Management of Data*, Houston, USA, 2018. Acceptance Rate: ∼20%.
54. A. Kim[⋆], L. Xu[⋆], T. Siddiqui[⋆], S. Huang[⋆], S. Madden, **A. Parameswaran**. Optimally Leveraging Density and Locality For Exploratory Browsing and Sampling. *HILDA '18: Workshop on Human-in-the-Loop Data Analytics* at the ACM SIGMOD Int'l Conf. on Management of Data, Houston, USA, 2018.
53. M. Bendre[⋆], V. Venkataraman[⋆], X. Zhou[⋆], K. Chang, **A. Parameswaran**. Towards a Holistic Integration of Spreadsheets with Databases: A Scalable Storage Engine for Presentational Data Management. *ICDE '18: 34th Int'l Conf. on Data Engineering*, Paris, France, 2018.
52. D. Lee[⋆], **A. Parameswaran**. The Case for a Visual Discovery Assistant: A Holistic Solution for Accelerating Visual Data Exploration, *IEEE Data Engineering Bulletin*, Special Issue on Insights and Explanations in Data Analysis, September 2018.
51. Y. Gao[⋆], **A. Parameswaran**, J. Peng. On the interpretability of conditional probability estimates in the agnostic setting *Electronic Journal of Statistics*, Volume 11, Number 2, January 2018. Extended Version of

AISTATS'17 paper.

50. K. Mack*, J. Lee*, K. Chang, K. Karahalios, **A. Parameswaran**, Characterizing Scalability Issues in Spreadsheet Software using Online Forums (Case Study). *CHI '18: Int'l Conf. on Human Factors in Computing Systems*, Montreal, Canada, 2018.
49. S. Rahman*, M. Aliakbarpour, H. Kong*, E. Blais, K. Karahalios, **A. Parameswaran**, and R. Rubinfeld. I've Seen Enough: Incrementally Improving Visualizations to Support Rapid Decision Making. *VLDB '17: 43rd Int'l Conf. on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ∼20%.
48. S. Huang*, L. Xu*, J. Liu*, A. Elmore, **A. Parameswaran**. OrpheusDB: Bolt-on Versioning for Relational Databases. *VLDB '17: 43rd Int'l Conf. on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ∼20%.. Invited to **Special Issue of VLDB Journal for VLDB 2017 Best Papers**.
47. A. Jain*, A. Das Sarma*, **A. Parameswaran**, and J. Widom. Understanding Workers, Developing Effective Tasks, and Enhancing Marketplace Dynamics: A Study of a Large Crowdsourcing Marketplace. *VLDB '17: 43rd Int'l Conf. on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ∼20%.
46. T. Siddiqui*, A. Kim*, J. Lee*, K. Karahalios, and **A. Parameswaran**. Effortless Visual Data Exploration with Zenvisage: An Expressive and Interactive Visual Analytics System. *VLDB '17: 43rd Int'l Conf. on Very Large Data Bases*, Munich, Germany, 2017. Acceptance Rate: ∼20%.
45. T. Rekatsinas, M. Joglekar*, H. Garcia-Molina, **A. Parameswaran**, and C. Re. SLiMFast: Guaranteed Results for Data Fusion and Source Reliability *SIGMOD '17: ACM SIGMOD Int'l Conf. on Management of Data*, Raleigh, USA, 2017 Acceptance Rate: ∼20%.
44. Y. Gao*, **A. Parameswaran**, and J. Peng. On the Interpretability of Conditional Probability Estimates in the Agnostic Setting (Oral Presentation). *AISTATS '17: Conf on Artificial Intelligence and Statistics*, Ft. Lauderdale, USA, 2017. Acceptance Rate: ∼30%. Invited to **Special Issue of EJS for AISTATS 2017 Best Papers**.
43. T. Siddiqui*, J. Lee*, A. Kim*, E. Xue*, X. Yu*, S. Zou*, L. Guo*, C. Liu*, C. Wang*, K. Karahalios, and **A. Parameswaran**. Fast-forwarding to Desired Visualizations with Zenvisage. *CIDR '17: Conf. on Innovative Data Management (CIDR)*, Chaminade, USA, 2017.
42. T. Siddiqui*, X. Ren, **A. Parameswaran**, and Jiawei Han. FacetGist: Collective Extraction of Document Facets in Large Technical Corpora, *CIKM '16: 25th Int'l Conf. on Information and Knowledge Management*, Indianapolis, USA, 2016. Acceptance Rate: 23%.
41. M. Maddox, D. Goehring, A. Elmore, S. Madden, **A. Parameswaran**, and A. Deshpande. Decibel: The Relational Dataset Branching System. *VLDB '16: 42nd Int'l Conf on Very Large Data Bases*, New Delhi, India, 2016. Acceptance Rate: ∼20%.
40. M. Vartak*, S. Rahman*, S. Madden, **A. Parameswaran**, and N. Polyzotis. SeeDB: Efficient Data-Driven Recommendations to Support Visual Analytics. *VLDB '16: 42nd Int'l Conf on Very Large Data Bases*, New Delhi, India, 2016. Acceptance Rate: ∼20%.
39. Y. Gao*and **A. Parameswaran**. Squish: Near-optimal Compression for Archival of Relational Datasets. *KDD '16: 22nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, San Francisco, USA, 2016. Acceptance Rate: 6%.
38. A. Das Sarma*, **A. Parameswaran**, and J. Widom. Towards Globally Optimal Crowdsourcing Quality Management. *SIGMOD '16: ACM SIGMOD Int'l Conf. on Management of Data*, San Francisco, USA, 2016 Acceptance Rate: 19%.
37. M. Joglekar*, H. Garcia-Molina, and **A. Parameswaran**. Interactive Data Exploration with Smart Drilldown. *ICDE '16: 32nd Int'l Conf on Data Engineering*, Helsinki, Finland, 2016. Acceptance Rate: 25%. Invited to **Special Issue of TKDE for ICDE 2016 Best Papers**.
36. M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *SIGMOD Record*, December 2016.
35. **A. Parameswaran**, A. Das Sarma*, and V. Venkataraman*. Optimizing Open-Ended Crowdsourcing: The Next Frontier in Crowdsourced Data Management. *IEEE Data Engineering Bulletin*, December 2016
34. H. Garcia-Molina, M. Joglekar*, A. Marcus, **A. Parameswaran**, and V. Verios. Challenges in Data Crowdsourcing. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, January 2016
33. A. Das Sarma*, A. Jain*, A. Nandi, **A. Parameswaran**, and J. Widom. Surpassing Humans and Computers with JellyBean: Crowd-Vision-Hybrid Counting Algorithms. *HCOMP '15: 3rd AAAI Int'l Conf. on Human*

Computation and Crowdsourcing, San Diego, USA, 2015.

32. S. Bhattacherjee, A. Chavan, S. Huang*, A. Deshpande, and **A. Parameswaran**. Principles of Dataset Versioning: Exploring the Recreation/Storage Tradeoff. *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.
31. A. Kim*, E. Blais, **A. Parameswaran**, P. Indyk, S. Madden, and R. Rubinfeld. Rapid Sampling for Visualizations with Ordering Guarantees. *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.
30. Y. Gao*and **A. Parameswaran**. Finish Them!: Pricing Algorithms for Human Computation, *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015. Acceptance Rate: 21%.
29. H. Zhuang*, **A. Parameswaran**, D. Roth, and J. Han. Debiasing Crowdsourced Batches, *KDD '15: 21st ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Sydney, Australia, 2015.Acceptance Rate: 20%.
28. A. Chavan, S. Huang*, A. Deshpande, A. Elmore, S. Madden, and **A. Parameswaran**. Towards a Unified Query Language for Provenance and Versioning, *TAPP'15: 7th Int'l Conf. on Theory and Practice of Provenance*, Edinburgh, Scotland, 2015.
27. M. Joglekar*, H. Garcia-Molina, **A. Parameswaran**, and C. Re. Exploiting Correlations for Expensive Predicate Evaluation, *SIGMOD '15: ACM SIGMOD Int'l Conf. on Management of Data*, Melbourne, Australia, 2015 Acceptance Rate: 26%.
26. M. Joglekar*, H. Garcia-Molina, and **A. Parameswaran**. Comprehensive and Reliable Crowd Assessment Algorithms, *ICDE '15: 31st Int'l Conf on Data Engineering*, Seoul, Korea, 2015. Acceptance Rate: 25%.
25. A. Bhardwaj, S. Bhattacherjee, A. Chavan, A Deshpande, A.J. Elmore, S. Madden, and **A. Parameswaran**. DataHub: Collaborative Data Science & Dataset Version Management at Scale, *CIDR '15: Conf. on Innovative Data Management (CIDR)*, Asilomar, USA, 2015.
24. **A. Parameswaran**, S. Boyd, H. Garcia-Molina, A. Gupta, N. Polyzotis, and J. Widom. Optimal Crowd-Powered Rating and Filtering Algorithms, *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014. Acceptance Rate: ∼20%.
23. A. Das Sarma, **A. Parameswaran**, H. Garcia-Molina, and A. Halevy. Crowd-Powered Find Algorithms, *ICDE '14: 30th Int'l Conf on Data Engineering*, Chicago, USA, April 2014. Acceptance Rate: 20%. Invited to **Special Issue of TKDE for ICDE 2014 Best Papers**.
22. **A. Parameswaran**, M. H. Teh, H. Garcia-Molina, and J. Widom. DataSift: An Expressive and Accurate Crowd-Powered Search Toolkit, *HCOMP '13: 1st AAAI Int'l Conf. on Human Computation and Crowdsourcing*, Palm Springs, USA, 2013. Acceptance Rate: 30%.
21. **A. Parameswaran**, R. Kaushik, and A. Arasu. Efficient Parsing-based Search over Databases, *CIKM '13: 22nd Int'l Conf on Information and Knowledge Management*, Burlingame, USA, 2013. Acceptance Rate: 16.8%.
20. M. Joglekar, H. Garcia-Molina, and **A. Parameswaran**. Evaluating the Crowd with Confidence, *KDD '13: 19th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Chicago, USA, 2013. Acceptance Rate: 17%.
19. K. Bellare, S. Iyengar, **A. Parameswaran**, and V. Rastogi. Active Sampling for Entity Matching with Guarantees, *ACM Transactions on Knowledge Discovery from Data*, Volume 7(3), September 2013. Extended Version of KDD'12 paper.
18. N. Dalvi, **A. Parameswaran**, and V. Rastogi. Minimizing Uncertainty in Pipelines, *NIPS '12: 25th Int'l Conf on Neural Information Processing Systems*, Tahoe, Nevada, USA, 2012.
17. **A. Parameswaran**, H. Park, H. Garcia-Molina, N. Polyzotis, and J. Widom. Deco: Declarative Crowdsourcing, *CIKM '12: 21st Int'l Conf on Information and Knowledge Management*, Maui, USA, 2012. Acceptance Rate: 13.4%.
16. K. Bellare, S. Iyengar, **A. Parameswaran**, and V. Rastogi. Active Sampling for Entity Matching, *KDD '12: 18th ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, Beijing, China, 2012. Acceptance Rate: 18%. Invited to **Special Issue of TKDD for KDD 2012 Best Papers**.
15. S. Guo, **A. Parameswaran**, and H. Garcia-Molina. So Who Won? Dynamic Max Discovery with the Crowd, *SIGMOD '12: ACM SIGMOD Int'l Conf. on the Management of Data*, Scottsdale, USA, 2012. Acceptance Rate: 17%.
14. **A. Parameswaran**, H. Garcia-Molina, H. Park, N. Polyzotis, A. Ramesh, and J. Widom. CrowdScreen:

Algorithms for Filtering Data with Humans *SIGMOD '12: ACM SIGMOD Int'l Conf. on the Management of Data*, Scottsdale, USA, 2012. Acceptance Rate: 17%.

13. F. Afrati, A. Das Sarma, D. Menestrina, **A. Parameswaran**, and J. D. Ullman. Fuzzy joins using MapReduce, *ICDE '12: 28th Int'l Conf. on Data Engineering*, Washington DC, USA, 2012. Acceptance Rate: 24%.
12. H. Park, R. Pang, **A. Parameswaran**, H. Garcia-Molina, N. Polyzotis, and J. Widom. An Overview of the Deco System: Data Model and Query Language; Query Processing and Optimization, *SIGMOD Record*, Volume 41, December 2012.
11. **A. Parameswaran**, N. Dalvi, H. Garcia-Molina, and R. Rastogi. Optimal Schemes for Robust Web Extraction, *VLDB '11: 37th Int'l Conf. on Very Large Data Bases*, Seattle, USA, 2011. Acceptance Rate: 18.1%.
10. **A. Parameswaran**, A. Das Sarma, H. Garcia-Molina, N. Polyzotis, and J. Widom. Human-assisted Graph Search: It's Okay to Ask Questions, *VLDB '11: 37th Int'l Conf. on Very Large Data Bases*, Seattle, USA, 2011. Acceptance Rate: 18.1%.
9. **A. Parameswaran** and N. Polyzotis. Answering Queries using Databases, Humans, and Algorithms, *CIDR '11: Conf. on Innovative Data Management (CIDR)*, Asilomar, USA, 2011.
8. G. Koutrika, H. Garcia-Molina, and **A. Parameswaran**. Information Seeking: Convergence of Search, Recommendations, and Advertising, *Communications of the ACM (CACM)*, November 2011.
7. **A. Parameswaran**, P. Venetis, and H. Garcia-Molina. Recommendation Systems with Complex Constraints: A Course Recommendation Perspective, *ACM Transactions on Information Systems (TOIS)*, Volume 29(4), November 2011.
6. **A. Parameswaran**, H. Garcia-Molina, and J. D. Ullman. Evaluating, Combining, and Generalizing Recommendations with Prerequisites, *CIKM '10: 19th Int'l Conf. on Information and Knowledge Management*, Toronto, Canada, 2010. Acceptance Rate: 13%.
5. **A. Parameswaran**, H. Garcia-Molina, and A. Rajaraman. Towards the Web of Concepts: Extracting Concepts from Large Datasets, *VLDB '10: 36th Int'l Conf. on Very Large Data Bases*, Singapore, 2010. Acceptance Rate: 18.4%. Invited to **Special Issue of VLDB Journal for VLDB 2010 Best Papers**.
4. **A. Parameswaran**, G. Koutrika, B. Berkovitz, and H. Garcia-Molina. Recsplorer: Recommendation Algorithms Based on Precedence Mining, *SIGMOD '10: ACM SIGMOD Int'l Conf. on the Management of Data*, Indianapolis, USA, 2010. Acceptance Rate: 21%.
3. A. Das Sarma, **A. Parameswaran**, H. Garcia-Molina, and J. Widom. Synthesizing View Definitions from Data, *ICDT '10: 13th Int'l Conf. on Database Theory*, Lausanne, Switzerland, 2010. Acceptance Rate: 36%.
2. B. Berkovitz, F. Kaliszan, G. Koutrika, H. Liou, **A. Parameswaran**, P. Venetis, Z. Zadeh, and H. Garcia-Molina. Social Sites Research Through CourseRank, *SIGMOD Record*, Volume 38, December 2009.
1. F. Cazals, **A. Parameswaran**, and S. Pion. Robust Construction of the Three-dimensional Flow Complex, *SOCG '08: ACM Symposium on Computational Geometry*, Maryland, USA, 2008. Acceptance Rate: 32%.

◇ **Books**

1. A. Marcus and **A. Parameswaran**. Crowdsourced Data Management: Industry and Academic Perspectives. *Foundations and Trends in Databases Series*, Vol. 6: No. 1-2, pp 1-161, Now Publishers, December 2015

◇ **Refereed Short, Demo, and Vision Papers**

27. X. Ji[⋆], **A. Parameswaran**, M. Hulsebos. Benchmarking Table Retrieval for Generative Tasks, TRL @ NeurIPS. *NeurIPS'24: 38th Int'l Conf on Neural Information Processing Systems*, Vancouver, 2024.
26. P. Kallanagoudar, C. Anand, R. Garcia, R. M. M. Hicke, **A. Parameswaran**, E. Jun, S. E. Chasins: Quilt: Custom UIs for Linking Unstructured Documents to Structured Datasets (Demo). *UIST'24: Int'l Conf on User Interface Software and Technology*, Pittsburgh, USA, 2024.
25. S. Shankar[⋆], **Aditya G. Parameswaran**. Building Reactive Large Language Model Pipelines with Motion (Demo). *SIGMOD'24: ACM Intl. Conf. on Management of Data*, Santiago, Chile, 2024.
24. F. Chen[⋆], D. Tang[⋆], H. Li[⋆], **A. Parameswaran**. Visualizing Spreadsheet Formula Graphs Compactly (Demo) *VLDB'23*: 49th International Conference on Very Large Data Bases, Vancouver, Canada. September 2023.
23. S. Shankar[⋆], B. Herman, **A. Parameswaran**. Rethinking Streaming Machine Learning Evaluation (Poster). ML Evaluation Standards Workshop at *ICLR'22:* Eleventh International Conference on Learning Representations, 2022.

22. D. Lee*, J. Lee*, R. Wang*, **A. Parameswaran**. ScatterSearch: Visual Querying of Scatterplot Visualizations (Poster). *VIS'19:* IEEE Int'l Conf. on Visualization, Vancouver, Canada, 2019.
21. M. Bendre*, T. Wattanawaroon*, S. Rahman*, K. Mack*, Y. Liu*, S. Zhu*, Y. Lu*, P. Yang*, X. Zhou*, K. Chang, K. Karahalios, **A. Parameswaran**. Faster, Higher, Stronger: Redesigning Spreadsheets for Scale. *ICDE'19: 35th Int'l Conf on Information and Knowledge Management*, Macau, 2019. **Best Demo Award (Awarded to 2 out of 24 demos)**.
20. S. Rajpal*, **A. Parameswaran**. Holistic Crowd-Powered Sorting via AID: Optimizing for Accuracies, Inconsistencies, and Difficulties (Short Paper), *CIKM'18: 27th Int'l Conf on Information and Knowledge Management*, Lingotto, Italy, 2018.
19. D. Xin*, L. Ma*, J. Liu*, S. Macke*, S. Song*, **A. Parameswaran**. Helix: Accelerating Human-in-the-loop Machine Learning (Demo). *VLDB'18: 44th Int'l Conf on Very Large Data Bases*, Rio De Janeiro, Brazil, 2018.
18. T. Siddiqui*, P. Luh*, Z. Wang*, K.Karahalios, **A. Parameswaran**. ShapeSearch: Flexible Pattern-based Querying of Trend Line Visualizations (Demo). *VLDB'18: 44th Int'l Conf on Very Large Data Bases*, Rio De Janeiro, Brazil, 2018.
17. D. Xin*, L. Ma*, S. Song*, **A. Parameswaran**. How Developers Iterate on Machine Learning Workflows — A Survey of the Applied Machine Learning Literature (Short Paper). *IDEA'18: Workshop on Interactive Data Exploration and Analysis* at the 24th KDD Int'l Conf. on Knowledge Discovery and Data Mining, London, UK, 2018.
16. D. Lee*, A. Das Sarma*, **A. Parameswaran**. Aggregating Crowdsourced Image Segmentations (Short Paper). *HCOMP '18: 6th Int'l Conf. on Human Computation and Crowdsourcing*), Zurich, Switzerland, 2018.
15. G. Su Yilmaz, T. Wattanawaroon*, L. Xu*, A. Nigam*, A. Elmore, **A Parameswaran**. DataDiff: User-Interpretable Data Transformation Summaries for Collaborative Data Analysis (Demo). *SIGMOD '18: ACM SIGMOD Int'l Conf. on Management of Data*, Houston, USA, 2018.
14. L. Xu*, S.Huang*, S. Hui*, A. Elmore, **A. Parameswaran**. OrpheusDB: A Light-weight Approach to Relational Dataset Versioning (Demo). *SIGMOD '17: ACM SIGMOD Int'l Conf. on Management of Data*, Raleigh, USA, 2017. **Best Demo Honorable Mention**.
13. M. Bendre*, B. Sun*, X. Zhou*, D. Zhang*, S. Lin*, K. Chang, and **A. Parameswaran**. Data-Spread: Unifying Databases and Spreadsheets (Demo) *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
12. M. Joglekar*, H. Garcia-Molina, and **A. Parameswaran**. Smart Drill-down: A New Data Exploration Operator (Demo). *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
11. A. Bhardwaj, A. Deshpande, A. Elmore, D. Karger, S. Madden, **A. Parameswaran**, H. Subramanyam, E. Wu, and R. Zhang. Collaborative Data Analytics with Datahub (Demo). *VLDB '15: 41st Int'l Conf on Very Large Data Bases*, Hawaii, USA, 2015.
10. S. Koltani, S. Wang, **A. Parameswaran**. GeoHashViz: Interactive Analytics for Mapping Spatiotemporal Diffusion of Twitter Hashtags (Poster). *XSEDE '15*, USA, 2015.
9. A. Das Sarma*, **A. Parameswaran**, and J. Widom. Optimal Worker Quality and Answer Estimates in Crowd-Powered Filtering and Rating (Short Paper). *HCOMP '14: 2nd Int' Conf. on Human Computation and Crowdsourcing*, Pittsburgh, USA, 2014.
8. M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *DSIA '15: Workshop on Interactive Analysis*, Chicago, 2015.
7. **A. Parameswaran**, N. Polyzotis, and H. Garcia-Molina. SeeDB: Visualizing Database Queries Efficiently (Vision), *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014.
6. M. Vartak*, S. Madden, **A. Parameswaran**, and N. Polyzotis. SeeDB: Automatically Generating Query Visualizations (Demo), *VLDB '14: 40th Int'l Conf on Very Large Data Bases*, Hangzhou, China, 2014.
5. **A. Parameswaran**, M. H. Teh, H. Garcia-Molina, and J. Widom. DataSift: A Crowd-Powered Search Toolkit (Demo), *SIGMOD '14: ACM SIGMOD Int'l Conf. on the Management of Data*, Snowbird, USA, 2014.

4. H. Park, R. Pang, **A. Parameswaran**, H. Garcia-Molina, N. Polyzotis, and J. Widom. Deco: A System for Declarative Crowdsourcing (Demo), *VLDB '12: 38th Int'l Conf on Very Large Data Bases*, Istanbul, Turkey, 2012.
3. S. Guo, **A. Parameswaran**, and H. Garcia-Molina. So Who Won? Dynamic Max Discovery with the Crowd (Poster), *CrowdConf '11: 2nd Crowdsourcing Conference*, San Francisco, USA, 2011.
2. **A. Parameswaran** and H. Garcia-Molina. Recommendations with Prerequisites (Short Paper) *RecSys '09: 3rd ACM Conf. on Recommender Systems*, New York, USA, 2009. Acceptance Rate: 43%.
1. E. Sadikov, **A. Parameswaran**, and P. Venetis. Blogs as Predictors of Movie Success, *ICWSM '09: AAAI Conf. on Weblogs and Social Media*, San Jose, USA, 2009.

◇ **Patents**

3. B. Atur, H. Elmeleegy, J. Lin Lee, **A. Parameswaran**, D. Petersohn, M. Vashishtha. Providing access to information in a relational database. Awarded 2024
2. B. Atur, H. Elmeleegy, J. Lin Lee, **A. Parameswaran**, D. Petersohn, M. Vashishtha. Systems and methods for converting dataframes to relational databases and/or vice versa. Awarded 2024
1. **A. Parameswaran**, S. Macke, D. Xin. Holistic optimization for accelerating iterative machine learning Awarded 2019.

◇ **Preprints**

10. S. Agarwal*, A. Gupta, M. Heddes, D. Soylu, R. Meng, P. T. J. Kon, J. Chen, M. Cemri, S. Liu, A. Gonzalez, A. Jain, B. Dalvi Mishra, L. Song, V. Meshram, X. Tang, J. Yoon, **A. Parameswaran**, I. Stoica, V. Mirrokni, B. Gokturk, P. Ranganathan, T. Pfister, N. Peng, C. Li. DuoDiscover: Scaling Autonomous Discovery Beyond the Evaluation Bottleneck. Technical Report. October 2026.
9. S. W. W. Yoo, C. Zeng, A. Somayajula, S. Ankireddy, R. Kosalge*, A. Khare, J. Mitchell, **A. Parameswaran**, R. Samdani, A. Datta. Meta-Reflective Evolution Effectively Improves Production AI Systems. Technical Report. October 2026.
8. Y. Lin*, C. Hao*, S. Shankar*, **A. Parameswaran**. Scout: Scalable Document Extraction via Data Similarity. Technical Report. August 2026.
7. S. Liu, A. Krentsel, S. Agarwal*, M. Cemri, Z. Mao, S. Ponnapalli, A. G. Dimakis, S. Ratnasamy, M. Zaharia, **A. Parameswaran**, I. Stoica. The Time is Here for Just-in-Time Systems: Challenges and Opportunities. Technical Report. May 2026.
6. A. Ailamaki, ..., **A. Parameswaran**, ... The Cambridge Report on Database Research. Technical Report. April 2025.
5. S. Chasins, A. Cheung, N. Crooks, A. Ghodsi, K. Goldberg, J. Gonzalez, J. Hellerstein, A. Joseph, M. Mahoney, **A. Parameswaran**, D. Patterson, R. Popa, K. Sen, S. Shenker, D. Song, I. Stoica. The Sky Above The Clouds. Technical Report. May 2022.
4. M. Bendre*, S. Rahman*, T. Wattanawaroon*, K. Mack*, Y. Lu*, K. Chang, K. Karahalios, **A. Parameswaran**. Directed Data Management: A New Frontier in Database Usability. Technical Report. August 2018
3. T. Wattanawaroon*, S. Macke*, **A. Parameswaran**. Towards a Theory of Data-Diff: Optimal Synthesis of Succinct Data Modification Scripts. Technical Report. December 2017
2. A. Jain*, K. Goel*, J. Young Seo*, A.Kuznetsov*, **A. Parameswaran**, H. Sundaram. It's a Matter of Perspective(s): Crowd-Powered Consensus Organization of Corpora. Technical Report. November 2016
1. L. Battle, T. Benson, **A. Parameswaran**, and E. Wu. Indexing Cost-Sensitive Prediction. Technical Report. August 2014

◇ **Other Writings**

27. S. Zeighami*, K. Akillioglu*, A. Biswal, A. Cheung, J. Gonzalez, **A. Parameswaran**. The Actor-Gatherer Framework for AI Agents. **EPIC Data Lab Blog**, October 2026.

26. **A. Parameswaran**, S. Agarwal*, ... Intelligence is Free, Now What? **BAIR Blog**, July 2026.
25. **A. Parameswaran**. Ten Papers from my First Decade as Faculty. **EPIC Data Lab Blog**, September 2025.
24. S. Zeighami*, S. Shankar*, **A. Parameswaran**. Low-Cost LLM-Powered Data Processing with BARGAIN. **EPIC Data Lab Blog**, September 2025.
23. Y. Lin*, A. Cheung, **A. Parameswaran**. Liberating Structured Data from PDF Prisons. **EPIC Data Lab Blog**, April 2025.
22. S. Shankar*, B. Chopra*, **A. Parameswaran**, E. Wu. Interactive LLM-Powered Data Processing with DocWrangler. **EPIC Data Lab Blog**, January 2025.
21. S. Amer-Yahia, L. Battle, Y. Hu, D. Moritz, **A. Parameswaran**, N. Bikakis, P. K. Chrysanthis, G. Li, G. Papastefanatos, L. Yu. Data Exploration and Visual Analytics Challenges in AI Era. **SIGMOD Blog**, January 2025.
20. **A. Parameswaran**. Big Data is Dead... Long Live Big Data. **Ponder Blog**, March 2023.
19. **A. Parameswaran**. Pandas vs. SQL – Part 4: Pandas Is More Convenient. **Ponder Blog**, August 2022.
18. **A. Parameswaran**. Pandas vs. SQL – Part 3: Pandas Is More Flexible. **Ponder Blog**, July 2022.
17. M. Vashishtha and **A. Parameswaran**. Pandas vs. SQL – Part 2: Pandas Is More Concise. **Ponder Blog**, June 2022.
16. **A. Parameswaran**. Pandas vs. SQL – Part 1: The Food Court and the Michelin-Style Restaurant. **Ponder Blog**, June 2022.
15. D. Lee, **A. Parameswaran**. Using Lux to visualize your pandas dataframes with zero effort. **Ponder Blog**, March 2022.
14. D. Petersohn, **A. Parameswaran**. How we parallelized 600+ pandas functions with Modin. **Ponder Blog**, March 2022.
13. **A. Parameswaran**. Visual Data Exploration: A Fertile Ground for Data Management Research. **SIGMOD Blog**, June 2018.
12. D. Lee*, T. Siddiqui*, K. Karahalios, **A. Parameswaran**. Effortless Exploration with Zenvisage 0.2, *Medium*, November 2017.
11. **A. Parameswaran**. Enabling Data Science for the Majority, **O'Reilly Blog**, October 2017.
10. A. Das Sarma*, **A. Parameswaran**. Crowdsourcing in Practice: Our Findings, *Medium*, September 2017.
9. S. Huang*, L. Xu*, A. Elmore, **A. Parameswaran**. Painless Data Versioning for Collaborative Data Science, *Medium*, September 2017.
8. S. Rahman*and **A. Parameswaran**. Draw Conclusions Early with Incvisage, *Medium*, September 2017.
7. T. Siddiqui*and **A. Parameswaran**. Towards Automating Insight, *Medium*, September 2017.
6. **A. Parameswaran**. Three Tools for Human-in-the-loop Data Analytics, *Medium*, January 2017.
5. M. Vartak*and **A. Parameswaran**. Instantly Visualizing What's Different, *Medium*, September 2016.
4. Y. Gao*and **A. Parameswaran**. On Optimally Squishing Large Datasets, *Medium*, August 2016.
3. A. Das Sarma*and **A. Parameswaran**. Revisiting Quality Management for Crowdsourcing: A Fresh Take, *Medium*, July 2016.
2. **A. Parameswaran**. Data Scavengers, not Research Parasites, *Medium*, January 2016.
1. A. Marcus, **A. Parameswaran**. The Fourth Industrial Revolution, *Medium*, January 2016.

Major Software Releases

◇ **DocETL**: An declarative framework for processing documents, with an agentic query optimizer, and an IDE DocWrangler. **140K+ Downloads; 4.1K+ GitHub Stars** as of October 2026, with usage across sectors. Deployed DocWrangler has been used 1000s of times, including by 30+ companies in the S&P 500. 2024– `https://github.com/ucbepic/docetl`

- ◇ **Actor-Gatherer**: A bolt-on framework for agent context management that decouples acting from proactive context gathering, working with Claude Code, Codex, Pi, and OpenCode. 2026– `https://github.com/ucbepic/actor-gatherer`
- ◇ **PRAP** (Police Records Access Project): A state-wide police records database, in collaboration with journalist and public defenders. Published simultaneously by LA Times, SF Chronicle, CalMatters, and KQED, enabling anyone to search through over 1M pages of PDFs from 500+ agencies in California. **Millions of visits**, and extensive usage by journalists and public defender groups. Led to new transparency legislation in October 2025. 2021– `https://bids.berkeley.edu/california-police-records-access-project`
- ◇ **Data Agent Benchmark**: A benchmark for evaluating data agents, developed in collaboration with PromptQL. Features a public leaderboard with dozens of entries. **194 GitHub Stars, 47 submissions** as of October 2026. 2025– `https://github.com/ucbepic/DataAgentBench`
- ◇ **TWIX**: A data extraction framework for templatized PDF documents. **230+ GitHub Stars** as of October 2026. 2025– `https://github.com/ucbepic/twix`
- ◇ **NUDGE**: Lightweight non-parametric fine-tuning of embeddings for retrieval. Deployed by LlamaIndex as part of their open-source framework. 2024– `https://github.com/szeighami/nudge`
- ◇ **BARGAIN**: A framework for low-cost LLM-powered data processing with statistical guarantees. Acknowledged in a Google Cloud blog post. 2025– `https://github.com/ucbepic/BARGAIN`
- ◇ **MLTrace**: A bolt-on ML pipeline monitoring and observability solution. **29K Downloads, 470 GitHub stars** as of October 2026. 2021–. `https://github.com/loglabs/mltrace`
- ◇ **Modin**: A scalable dataframe system. Modin uses database system techniques to speed up the execution of Pandas API commands, while acting as a drop-in replacement for Pandas. **68M Downloads; 10.4K+ GitHub Stars** as of October 2026. 2020–. `https://github.com/modin-project/modin`
- ◇ **Nbsafety** (now called Ipyflow): Nbsafety adds a layer of protection to computational notebooks to prevent errors and non-reproducible behavior. Nbsafety eliminates stale dependencies by providing warnings to users and is a simple extension to Jupyter Notebooks. **570K Downloads, 1.3K GitHub Stars** as of October 2026. 2020–. `https://github.com/ipyflow/ipyflow`
- ◇ **Lux**: A python API for intelligent visual discovery. Lux acts as an alternative visualization for a dataframe, providing automated recommendations of next steps in analysis. **700K+ Downloads, 5.4K GitHub Stars** as of October 2026. 2020–. `https://github.com/lux-org/lux`
- ◇ **TARGET**: Table Retrieval Benchmark. 2024–. `https://github.com/target-benchmark/target`
- ◇ **Covidvis**: Visualizing the impact of various forms of interventions on the progress of SARS-COV-2 across various states and countries. Developed in collaboration with epidemiologists and public health experts. 2020–. `http://covidvis.github.io`
- ◇ **Spreadsheet Benchmark**: A performance benchmark for three spreadsheet systems: Google Sheets, Libreoffice Calc, and Microsoft Excel, testing the latency of various operations as well as the presence of optimizations. 2020–. `https://github.com/dataspread/spreadsheet-benchmark`
- ◇ **DataSpread**: A Spreadsheet-Database Hybrid. DataSpread has a spreadsheet frontend, and a database backend. DataSpread inherits the flexibility and ease-of-use of spreadsheets, as well as the scalability and power of databases, and scales to billions of cells seamlessly. 2016–. `http://dataspread.github.io`
- ◇ **OrpheusDB**: A relational dataset versioning system. OrpheusDB is built on top of standard relational databases, thus it inherits much of the same benefits of relational databases, while also compactly storing, tracking, and recreating versions on demand, all very efficiently. 2016–2020. `http://orpheus-db.github.io`
- ◇ **Zenvisage**: An 'effortless' data visualization tool. Zenvisage can automatically identify and recommend visualizations that match desired user patterns. The user can specify at a high level what they are looking for either via interactions or via a query language (ZQL), and the system will perform the necessary computation to identify these visualizations. 2016–2019. `http://zenvisage.github.io`

Funding

- ◇ Principal Investigator, **NSF-STRIDE AI Efficiency Challenge Award**, for improving the cost and quality of large-scale LLM-powered data processing, one of nine teams selected nationwide, 2026-2028.
- ◇ Principal Investigator, **Laude Slingshot Grant**, 2026.
- ◇ Principal Investigator, **State of California** Police Records Project, 2023-2026.
- ◇ Co-Principal Investigator, **NSF NRT Award** titled "CRELS: Computational Research for Equity in the Legal System Training Program", 2023-2028.
- ◇ Principal Investigator, **NSF NAIRR Pilot Credit Award**, 2025-2026.
- ◇ Co-Director, **EPIC** (Effective Programming, Interaction, and Computation with) **Data Lab**, supported via industry sponsors (including Adobe, Bridgewater, Google, Jane Street, Microsoft, PromptQL, Sigma Computing, and Snowflake). 2022-2027.
- ◇ Principal Investigator, **NSF FoW Award** titled "Human-Machine Teaming for Effective Data Work at Scale: Upskilling Defense Lawyers Working with Police and Court Process Data", 2021-2024.
- ◇ Principal Investigator, **Sloan Fellowship Award**, 2020-2022.
- ◇ Principal Investigator, **Army Research Office Young Investigator Program** Award, 2018–2021.
- ◇ Principal Investigator, Toyota Research Institute **Accelerated Materials Design and Discovery** Grant, Jointly Awarded to Carnegie Mellon University and Illinois, 2018–2019.
- ◇ Principal Investigator, **NSF AITF Award** titled "Fast, Accurate, and Practical: Adaptive Sublinear Algorithms for Scalable Visualization", Jointly awarded to MIT, USC, and Illinois, 2017–2020.
- ◇ Principal Investigator, **NSF CAREER Award** titled "Advancing Open-Ended Crowdsourcing: The Next Frontier in Crowdsourced Data Management", 2017–2021.
- ◇ Amazon Web Services **Research Credits**, 2017.
- ◇ Adobe Research **Faculty Research Gift**, 2016.
- ◇ Principal Investigator, **Google Focused Research Award**, titled "Holistic Optimization for Accelerating Iterative Machine Learning", 2017.
- ◇ Co-Principal Investigator, **NSF BIGDATA Award** titled "Bringing Interactive Data Management to Scientists, Analysts, and the Masses: A Holistic Unification of Spreadsheets and Databases", 2016–2020.
- ◇ Principal Investigator, **Siebel Energy Institute** Seed Grant, "Data-driven Discovery of Resilient Energy Storage for Grid Applications", Jointly awarded to Carnegie Mellon University and Illinois, 2016–17.
- ◇ Principal Investigator, **NIH BD2K** (Big Data to Knowledge) Supplementary Grant, "Piloting a Data Publication Service for the BD2K Commons", Jointly awarded to U Chicago and Illinois, 2015–16.
- ◇ Principal Investigator, **NSF IIS Medium Award** titled "DataHub: Collaborative Dataset Management for Data Science", Jointly awarded to Illinois, Maryland, and MIT, 2015–2018.
- ◇ Google **Faculty Research Award**, 2015.
- ◇ Investigator, **NIH BD2K** (Big Data to Knowledge) Centers of Excellence titled "KnowEng, a Scalable Knowledge Engine for Large Scale Genomic Data", Jointly awarded to Illinois and Mayo Clinic, 2014–2018.
- ◇ Google Cloud Credit Award, 2014.

Industry Internships

- ◇ **Yahoo! Research**, Web Information Management Group, Santa Clara, CA — Summer 2011
- ◇ **Microsoft Research**, Data Management, Mining and Exploration Group, Redmond, WA — Summer 2010
- ◇ **Yahoo! Research**, Information Extraction Group, Bangalore, India — Winter 2009
- ◇ **Kosmix**, Mountain View, CA — Summer 2008
- ◇ **INRIA**, Geometric Algorithms Group, Sophia-Antipolis, France — Summer 2006
- ◇ **Microsoft Research**, Formal Methods Group, Bangalore, India — Winter 2005

Mentorship

- ◇ Graduated PhD Students Supervised [All students at Berkeley unless indicated otherwise.]

- Shreya Shankar,
  Title: TBD (Completing Fall 2026)
  2021—2026 (Winner: **Heidelberg Laureate Fellowship**, Hertz Fellowship Finalist, Soros Fellowship Finalist, **NDSEG Fellowship**, Apple AI/ML Scholar Campus Nominee, EECS Rising Stars Awardee)
  First employment: **Assistant Professor, CMU CS Department**, starting Fall 2027;
- Doris Lee,
  Title: "*Designing Automated Assistants for Visual Data Exploration*"
  2017—2021 (Winner: **Forbes 30 under 30**, **CHI Dissertation Consortium**, **Facebook Fellowship**, awarded to 40 out of 1600 applicants, **UIUC Computer Science Excellence Fellowship**, IUI Travel Award 2019, CRA-W Travel Award 2019)
  First employment: CEO, Ponder (Acq. Snowflake).
- Doris Xin,
  Title: "*Usable and Efficient Systems for Machine Learning*"
  2016—2021 (Winner: **NSF GRFP**, **Heidelberg Laureate Fellowship**, SIGMOD Travel Award 2018, 2017)
  First employment: CEO, Disarray.
- Devin Petersohn,
  Title: "*Dataframe Systems: Theory, Architecture, and Implementation*"
  2015—2021 (Winner: **NSF GRFP**, **James Tullock Memorial Scholarship Award**)
  First employment: CTO, Ponder (Acq. Snowflake).
- Illinois: Stephen Macke,
  Title: "*Leveraging Distributional Context for Safe and Interactive Data Science at Scale*"
  2015—2021 (Winner: **NSF GRFP**, **State Farm Doctoral Scholars Fellowship**)
  First employment: Research Scientist at Meta → Databricks.
- Illinois: Tana Wattanawaroon,
  Title: "*Generalizing Spreadsheet Computation for Evolving Spreadsheets at Scale*"
  2015—2021 (Winner: **Outstanding TA Award**, SIGMOD Travel Award 2017, ICDE Travel Award 2019, SIGMOD Travel Award 2019)
  First employment: Faculty, Kasetsart University.
- Illinois: Tarique Siddiqui,
  Title: "*Towards Expressive and Scalable Visual Data Exploration*"
  2016–2020 (Winner: **Siebel Scholarship**, VLDB Travel Award 2018)
  First employment: Senior Researcher at Microsoft Research.
- Illinois: Sajjadur Rahman,
  Title: "*Intuitive, interactive, and scalable multi-resolution interfaces for accelerating data exploration*"
  2014–2020 (Winner: ICDE Travel Award 2019)
  First employment: Senior Researcher at Megagon Labs.
- Illinois: Liqi Xu,
  Title: "*New Capabilities for Large-scale Exploratory Data Analysis*"
  2015–2020
  First employment: Research Scientist at Meta.
- Illinois: Silu Huang,
  Title: "*Effective Data Versioning for Collaborative Data Analytics*"
  (Awarded **ACM SIGMOD Jim Gray Dissertation Award Honorable Mention**),
  2014–2019 (Winner: **3M Fellowship**, 2015, **MSR Fellowship**, 2017—the first UIUC winner in 7 years)
  First employment: Senior Researcher at Microsoft Research → Bytedance
- Illinois: Mangesh Bendre,
  Title: "*Towards Unifying Spreadsheets with Databases for Ad-hoc Interactive Data Management at Scale*",
  2014-2018 (Co-advisor: Kevin Chang)
  First employment: Research Scientist at Visa Research.

- Illinois: Yihan Gao,
  Title: "*Extracting and Utilizing Hidden Structures in Large Datasets*",
  2013–2019 (Winner: **Richard T. Cheng Fellowship**)
  First employment: Assistant Professor at Tsinghua University.
- Stanford: Akash Das Sarma,
  Title: "*Advancing the Use of Crowdsourcing for Data-intensive Tasks*",
  2012–2017 (Co-advisor: Jennifer Widom)
  First employment: Research Scientist at Facebook.

◇ Past Postdocs Advised

- Yiming Lin, 2023—2026, PhD from UCI,
  First employment: Assistant Professor at **NTU Singapore**, starting Fall 2026;
- Madelon Hulsebos, 2023—2024, PhD from UAmsterdam,
  Winner: BIDS Postdoctoral Fellowship
  First employment: Assistant Prof at **CWI Amsterdam**;
- Dixin Tang, 2021—2024, PhD from UChicago
  First employment: Assistant Prof at **UT Austin**;

◇ Current PhD Students or Postdocs

- Bhavya Chopra, 2024— (Winner: **Workday AI PhD Fellowship, Jane Street Fellowship Finalist**);
- Sepanta Zeighami, 2024—, Postdoc (PhD from USC)
- Shubham Agarwal, 2025— (Coadvisor: Ion Stoica);
- Ruiying Ma, 2025—;
- Lindsey Wei, 2026— (Coadvisor: Alvin Cheung);
- Daniel Lee, 2026— (Coadvisor: Natacha Crooks);

◇ MS Thesis Students

- Rachel Lin,
  Title: "*AI-Assisted Dataset Discovery with DATASCOUT*",
  2024–25 (First employment: Opto);
- Mawil Hasan,
  Title: "*Benchmarking Extraction of Structured Data from Templatized Documents*",
  2024–25 (First employment: Schonfeld);
- Todd Yu,
  Title: "*Efficient Visualization Recommendation under Updates*",
  2022–23 (First employment: Databricks);
- Joshua Wu,
  Title: "*Smash: A Dictionary-Free String Distance Metric that Considers Acronyms, Abbreviations, and Typos*",
  2022–23 (First employment: Plaid);
- Kunal Aggarwal,
  Title: "*Accelerating Visual Data Exploration via Sampling: A Case Study with Lux*",
  2021–22 (First employment: Stripe);
- Chris De Leon,
  Title: "*Towards Efficient Spreadsheet Computation and Visualization*",
  2021–22 (First employment: Google);
- Priyam Mohanty,
  Title: "*A Case Study on COVID-19 Intervention Visualizations: The Role of Trust, Beliefs, and Interpretations*",
  2021-22 (First employment: Google);
- Nidhi Kakulawaram,
  Title: "*Spreadsheet Bubbles: Showing Contextually Relevant Data During Formula Editing*",
  2020-21 (First employment: Twilio);

- Jerry Song,
  Title: "*Understanding and Evaluating the User Data Exploration Experience with Lux*",
  2020–21 (First employment: Goldman Sachs);
- Vincent Truong,
  Title: "*Generalized Partitioning for Dataset Versions in OrpheusDB*",
  2020–21 (First employment: Redfin);
- Avi Arjavalingam,
  Title: "*HASTE: Serverless DAG Execution Optimizer*",
  2020–21 (First employment: Abacus.ai);
- Richard Lin,
  Title: "*A Spreadsheet Interface for Dataframes*",
  2020–21 (First employment: Rockset);
- Illinois: Jaewooo Kim,
  Title: "*Assisting data exploration via in-situ adaptive visualizations*",
  2018–20 (First employment: Amazon);
- Illinois: Angela Lee,
  Title: "*Demystifying a dark art: Understanding real-world machine learning model development*",
  2019-20 (First employment: Google)
- Illinois: Renxuan Wang,
  Title: "*Enabling Efficient Visual Data Exploration for Solvent Discovery in Material Science*",
  2017–19 (First employment: Google);
- Illinois: Jialin Liu,
  Title: "*Machine Learning Workflow Optimization via Automatic Discovery of Resource Reuse Opportunities*",
  2017–19 (First employment: LinkedIn);
- Illinois: Zesheng Wang,
  Title: "*Efficient Pattern-based Querying of Trendline Visualizations*",
  2016—18 (First employment: Roblox);
- Illinois: Edward Xue,
  Title: "*Towards a Scripting Language for Visual Data Exploration*",
  2015—18 (First employment: Facebook);
- Illinois: Litian Ma,
  Title: "*Towards Understanding and Simplifying Human-in-the-loop Machine Learning*",
  2016—18 (First employment: Amazon);
- Illinois: Vipul Venkataraman,
  Title: "*DataSpread: scaling spreadsheets using relational databases*",
  2015—17 (Winner: **Siebel Scholarship**, **Outstanding TA Award**, First employment: Google);
- Illinois: Tarique Siddiqui,
  Title: "*Effortless data exploration with Zenvisage*",
  2014—16 (Winner: **Siebel Scholarship**, First employment: PhD student at Illinois);
- Illinois: Sili Hui,
  Title: "*OrpheusDB: an attempt at dataset version control within a relational database*",
  2014—16 (First employment: Medallia);
- Illinois: Ayush Jain,
  Title: "*Towards open-ended crowd-powered data processing: a case study of clustering and counting*",
  2014—16 (Winner: **Muroga Fellowship**, First employment: Google);

◇ Visiting Researchers

- Kerem Akillioglu (PhD student from Waterloo) 2026–;
- Haotian Li (PhD student from HKUST) 2022–23;
- Yong Wang (PhD student from Tsinghua) 2022–22;

◇ PhD/MS Students (Informal or Past advising)

- Wenjing Lin, 2023–24 (First employment: EPIC Software); Jennifer Momoh, 2020–21 (First employment: Linkedin); Thyne Boonmark, 2019–21 (First employment: Walmart Global Tech); Eva Wu, 2019–21 (First employment: PhD program at U Zurich);
- Illinois: Ti-Chung Cheng, 2018—20 (First employment: PhD program at U Illinois); Chaoran Wang, 2015—18 (First employment: Uber); Shreya Rajpal, 2016—18 (First employment: Drive.ai, now CEO and Co-Founder, Guardrails.ai); Himel Dev, 2015—16;
- MIT: Albert Kim, 2013—16; Manasi Vartak, 2013—15 (First employment: CEO and Founder, Verta);
- Stanford: Manas Joglekar, 2012—16 (First employment: Google); Ming Han Teh, 2012—14 (Winner: **Stanford's MS Thesis Award**, First employment: Microsoft); Aditya Ramesh, 2011—12 (First employment: LinkedIn);

◇ Undergraduate Students

- Current: Chiyu Hao, 2025–; Youran Sun, 2025–; Adit Chintamaneni, 2026–; Andrew Cheng, 2026–; Chanbin Park, 2026–; Sasha Singh, 2026–; Rohan Kosalge, 2024–;
- Past students opting for Ph.D. studies:
  - Lindsey Wei, 2025 (**CRA Outstanding Undergraduate Researcher Honorable Mention**; First employment: Berkeley; declined offers from MIT, CMU)
  - Harrison Huang, 2025 (First employment: ASU)
  - Ruiqi Chen, 2025 (First employment: Michigan)
  - Ruiying Ma, 2024–25 (First employment: Berkeley; declined offers from MIT, CMU)
  - Reya Vir, 2024–25 (**NSF GRFP**; First employment: Columbia)
  - Parth Asawa, 2022–24 (**Berkeley Campus Nominee for CRA Undergraduate Research Award**; First employment: Berkeley; declined offer from MIT, CMU)
  - Zixuan Yi, 2022–23 (First employment: U Penn; declined offers from Madison, Maryland)
  - Fanchao Chen, 2022–23 (First employment: Wisconsin; declined offers from UIUC, GAtech, Michigan, Penn, Purdue, and Maryland)
  - Rebecca Hicke, 2021—22 (First employment: Cornell; declined offers from UIUC)
  - Murtaza Ali, 2020—21 (First employment: UWashington HCDE)
  - William Ma, 2019—20 (First employment: UChicago)
  - Kelly Mack, 2015—19 (Winner: **NSF GRFP** 2019, **Snap Research Scholarship** 2017–awarded to 8 undergraduate/masters students, **CRA Undergraduate Research Award** Honorable mention 2017; Boeing Women in Engineering Scholarship 2018; NVIDIA Memorial Scholarship 2018; Silvio and Loretta Corsetti Scholarship 2017; First employment: UWashington);
  - Andrew Kuznetsov, 2014—18 (Winner: **Illinois SURF Research Award**, 3^{rd} place at the **Engineering Open House**; First employment: CMU);
  - Changfeng Liu, 2016 (UG Intern; First employment: UMichigan);
  - Yuxuan Zou, 2016–17 (First employment: Illinois → NUS);
  - Karan Goel, 2015–16 (UG Intern, First employment: CMU → Stanford);
  - Yinjun Wu, 2015-16 (UG Intern; First employment: UPenn);
- Past students opting for M.S. studies outside Berkeley: Ruilin Zhang, 2018–19 (First employment: USC); Jaewoo Kim, 2017—18 (First employment: UIllinois); Paul Luh, 2016—18 (First employment: UWisconsin); Jialin Liu, 2016 (UG Intern; First employment: UIllinois); Shreya Rajpal, 2015–16 (UG Intern, First employment: UIllinois); Yulun Du, 2014—15 (First employment: CMU); Ding Zhang, 2014—16 (First employment: CMU); Xinyan Zhou, 2014—16 (First employment: UIllinois);
- Past students opting for non-academic employment: Carl Ji, 2022–24 (First employment: Nooks); Pragya Kallanagoudar, 2022–24 (First employment: Headway); Quentin Romero Lauro, 2024–26 (Winner: **CRA Outstanding Undergraduate Researcher Finalist**; First employment: YC Founder); Yash Jain, 2024–25 (First employment: YC Founder); Yujie Wang, 2022–23 (First employment: Google); Avinash Rao, 2022–23 (First employment: Databricks); Christina Fan, 2022–23 (First employment: Amazon); Jonathan Yun, 2021–23 (First employment: Oracle); Bella Mendoza, 2022-2022 (First employment: OSTP, White House); Connor Lien, 2022–2023 (First employment: Databricks); Micah Yong, 2021–2022 (First employment: Applied Intuition); Aditi Mahajan, 2021—2022 (First employment: Google); Noah Kuo, 2020—21 (First employment: Google); Jared Zhao, 2020—21 (First employment: Cofounder, Polyture); Bangqi Wang, 2016—17

(First employment: NVidia); Abhishek Nigam, 2016—17 (First employment: Jump); Gary Luo, 2015—16 (First employment: Yahoo!); Joon Seo, 2014—16 (First employment: Google); Anurag Choudhury, 2015—16 (First employment: Microsoft)

PhD/MS Thesis Committees

◇ I have participated in the following thesis committees. Berkeley is the default, unless indicated otherwise.

- Chanwut Kittivorawong, 2026, Advisor: Alvin Cheung
- Asim Biswal, 2026, Advisor: Joseph Gonzalez
- Vinamra Benara, 2026, Advisor: Ion Stoica
- Xiaoyuan Liu, 2026, Advisor: Dawn Song
- Jichan Chung, 2025, Advisor: Irene Chen, Kannan Ramachandran
- David Chu, 2025, Advisor: Natacha Crooks, Joe Hellerstein
- Chase Stokes, 2025, Advisor: Marti Hearst
- Justin Lubin, 2025, Advisor: Sarah Chasins
- Parker Zeigler, 2025, Advisor: Sarah Chasins
- Columbia: Yiru Chen, 2024, Advisor: Eugene Wu
- CMU: Alex Cabrera, 2023, Advisor: Adam Perer, Jason Hong
- Duke: Zhengjie Miao, 2022, Advisor: Sudeepa Roy
- Samantha Robertson, 2022, Advisor: Niloufar Salehi (I am co-chair on paper.)
- Illinois: Kiumars Soltani, 2017, Advisor: Shaowen Wang
- Illinois: Fangbo Tao, 2016, Advisor: Jiawei Han
- Illinois: Yinan Zhang, 2016, Advisor: Cheng Zhai
- Illinois: Yodsawalai Chodpathumwan, 2016, Advisor: Marianne Winslett
- Illinois: Muntasir Rahman, 2015, Advisor: Indranil Gupta
- Illinois: Tanvir Amin, 2015, Advisor: Tarek Abdelzaher
- Illinois: Jialu Liu, 2015, Advisor: Jiawei Han

◇ I have served as a second signature for the following MS Theses: Erin Tan (2026), Jeremy Ferguson (2024), Roshan Regula (2024), Zoey Sun (2023), Dhanya Jayagopal (2022).

External Service Activities

◇ **Steering Committees:**

- DASHSys (Systems for Data-centric Agents with Human-in-the-loop) Workshop at VLDB, 2026.
- SAO (Supporting Our AI Overlords) Workshop on AI Agents and Data Systems at CAIS, 2026.
- Data AI Systems Workshop Steering Committee at ICDE, 2025.
- Human-in-the-Loop Data Analytics (HILDA) Workshop Steering Committee at SIGMOD, 2016—
- Data Systems for Interactive Analysis (DSIA) Workshop Steering Committee at VIS, 2018—

◇ **Chair/Editor Positions:**

- Area Chair, SIGMOD 2027
- Area Chair, SIGMOD 2026
- Associate Editor, VLDB 2025 (Demo).
- Associate Editor, VLDB Journal 2025—
- Chair, Workshops, SIGMOD 2021.
- Sponsor Chair (US), VLDB 2021.
- Chair, Workshops, SIGMOD 2020.
- Area Chair, SIGMOD 2020.
- Associate Chair, HCOMP 2019.
- Associate Editor, VLDB April 2019—March 2020.
- Associate Editor, SIGMOD Record, December 2014—September 2019.
  In charge of "Vision" articles, targeting visionary ideas and projects.
- Co-Chair, Human-in-the-loop Data Analytics Workshop (HILDA) at SIGMOD, 2017.
- Co-Chair, SIGMOD Undergraduate Research Competition, 2016.
- Area Chair, SIGMOD 2017

- Chair, Workshops and Tutorials, 1st HCOMP (Human Computation Conference) 2013
- Guest Editor, ACM XRDS Magazine (Issue on "Big Data"), with 15 invited articles from experts, 2012

◇ **Tutorials:**

- Invited full day Tutorial at the VLDB Summer School at Renmin University, Beijing, 2019.
- Invited 3-hour Tutorial at the HCOMP (Human Computation and Crowdsourcing) Conference 2016, titled *"Crowdsourced Data Management: Industry and Academic Perspectives"*, with Adam Marcus, 2016.

◇ **Working Groups:**

- Database Meeting in Boston
  Once-in-five-year convening of experts in data management to determine the future of the field, 2024.
- NIH BD2K Commons Working Group, 2015–18.
- NIH BD2K Machine Learning Working Group, 2016–18.

◇ **Award Committees:**

- CIDR Test of Time Award Committee, 2026–2027
- SIGMOD Jim Gray Dissertation Award Committee, 2015
- SIGMOD Best Demo Award Committee, 2014
- PURE Undergraduate Research Award Committee, 2014

◇ **Panel Participation:**

- Panel, NorCal DB Day, UC Santa Cruz, October 2026
- Panel, "Data Agents: Rethinking Data Systems in the AI Agent Era" at VLDB, August 2026
- Panel, DASHSys Workshop at VLDB, August 2026
- Panel, "Agents-meets-Databases" at ICDE, May 2026
- Panel Moderator, DEEM Workshop at SIGMOD, June 2026
- Panel, "LLM-meets-Databases" at SIGMOD, July 2025
- Faculty Panel on Agents, Snowflake Dev Day, June 2025
- Panel Discussion, "Pandas v. SQL" at Scalable Pandas Meetup, 2022
- Plenary Panel "DB meets AI" at the Alibaba Workshop@SIGMOD, 2018
- Plenary Panel "DB meets Vis" at HILDA@SIGMOD, 2018
- Plenary Panel "Will AI Eat Us All?" at VLDB, 2016
- Enterprise Intelligence Workshop Panel at KDD, 2016

◇ **Proposal Reviews:**

- Army Proposal Review, December 2022
- NSF IIS-III Databases Panel, Nov 2019
- NSF IIS-III Databases Panel, May 2019
- NSF IIS-III Databases Panel, April 2017
- NSF IIS-III Databases Panel, April 2016
- NSF IIS-III Data Mining Panel, June 2015

◇ **Program Committee Member:**

- ACL Rolling Review 2026
- ICDE 2027 DEFT
- SIGMOD 2026 (Industry Track)
- CIDR 2027, 2026, 2025
- VLDB 2027, 2024-25 (Rapid Response Reviewer), 2023-24
- VLDB 2025 (Tutorials)
- CIDR 2025
- ICDE 2024 (Demo)
- VLDB 2019 (Demo)
- ICDE 2018 (Demo)
- SIGMOD 2017 (Demo)

- HCOMP 2017, 2014, 2012
- VLDB 2016, 2015, 2014
- KDD 2015
- SIGMOD 2015, 2014
- SOCC 2014
- WSDM 2014
- ICDE 2014
- EDBT 2014
- WWW 2014

◇ **Workshop Program Committee:** HILDA 2019, 2018, 2016, IDEA 2015, SIGMOD ExploreDB 2015, COMAD 2014, CIKM CloudDB 2013, SIGMOD DBSocial 2013

◇ **Journal Reviewer:** VLDB J., 2015; J. Data & Info. Quality, 2014; SIGMOD Record, 2014; ACM TODS, 2014; Info. Sys., 2013-14; CACM, 2013; SIAM J. of Comp., 2013; IEEE TKDE, 2015-16, 2010-13; ACM TOIS, 2012; J. Web Semant., 2011

◇ **Workshop Participation:** Dagstuhl Workshop on "Managing Vector Data for Retrieval Augmented Generation: Systems and Algorithms", Dagstuhl, Germany, 2026; BIRS-CMO Workshop on "Theory and Models of Crowds and Networks", Oaxaca, Mexico, 2016

◇ **External Letters:** Tenure Evaluation: 2026 (university in China), 2026 (university in IL), 2026 (university in IL, Declined), 2025 (university in IL), 2025 (university in NY, Declined), 2024 (university in IL), 2023 (university in WA), 2019 (liberal arts college); Sloan Fellowship Recommendation: 2025, 2024 (2×)

Internal Service Activities

◇ CDSS/EECS Future Spaces Committee, 2025–26
◇ EECS Gateway 2nd Floor Space Coordinator, 2025–2026
◇ EECS Faculty Equity Advisor, (Regular CS Faculty and AIIS Search), 2025–26
◇ EECS Tenure Ad-Hoc Committee, 2025–26
◇ University Regents and Chancellors Review, 2025–26
◇ University Fiat Lux Review, 2025–26
◇ University External MIDS Class Evaluation, 2025–26
◇ University BRDO Schmidt Campus Committee, 2025–26
◇ EECS Faculty Equity Advisor (Regular CS Faculty Search) 2024–25
◇ EECS Faculty Equity Advisor (Simons Chair Search) 2024–25
◇ EECS Faculty Equity Advisor (AI, Inequality and Society Search) 2024–25
◇ EECS Gateway 2nd Floor Space Coordinator, 2024–2025
◇ CDSS/EECS Ad-Hoc Committee on Interdisciplinary Data Science Department 2024
◇ CDSS/EECS Future Spaces Committee 2024–25
◇ I School Ad Hoc Graduate Academic Misconduct Committees, 2023–24
◇ I School Ad Hoc Grievance Committee, 2023–24
◇ I School Equity Advisor, 2023
◇ I School Teaching Professor Evaluation Committee, 2023–24
◇ I School MIDS Executive Director Search Committee, 2023–24
◇ I School MIMS Program & Curriculum & Admissions, 2023–24
◇ I School Assistant Dean Search Committee, 2024
◇ VCRO Proposal Review Committee 2024
◇ EECS Ad Hoc Committee (Mid Career Review) 2023
◇ I School Adjunct Professor Search 2023
◇ I School Lecturer Assessment 2022
◇ I School Lecturer Excellence Review 2022
◇ EECS Graduate Admissions Committee 2019–21
◇ CDSS DataHub Space Committee 2020–21
◇ I School Faculty Equity Advisor 2021
◇ I School PhD Admissions 2021

- ◇ I School MIMS Admissions 2021
- ◇ I School MIDS Curriculum Committee 2019–20
- ◇ I School PhD Program Committee 2019–20
- ◇ Illinois, Big Research Initiatives Committee 2017–19
- ◇ Illinois, GEBI Chair Recruiting Committee 2017–19
- ◇ Illinois, Subcommittee for Review of New Courses 2017
- ◇ Illinois, Undergraduate Study Committee 2016–17
- ◇ Illinois, By Laws Committee 2015–17
- ◇ Illinois, Faculty Recruiting Committee 2014–15

Teaching

- ◇ **Instructor for DATA101/INFO258/CS187**: Data Engineering, Spring 2024, Spring 2025, Fall 2025, Fall 2026, Berkeley
  Students Enrolled: 75 (2024), 248 (Spring 2025), 395 (Fall 2025), 400 (Fall 2026)
- ◇ **Instructor for CS286**: Graduate Databases, Spring 2026, Berkeley
  Students Enrolled: 8 (2026)
- ◇ **Instructor for INFO 202**: Information Organization and Retrieval, Fall 2024, Berkeley
  Students Enrolled: 80 (2024)
- ◇ **Instructor for INFO290**: Human-Centered Data Management (Graduate Topics Class), Fall 2023, Berkeley
  Students Enrolled: 15 (2023)
- ◇ **Co-Instructor for CS194/290**: Data Engineering (Undergraduate/Graduate Class), Spring 2021, Berkeley
  Students Enrolled: 125 (2021); co-taught with Joe Hellerstein
  New class developed for the data science major. Now part of the core curriculum as DATA 101, to be taken after DATA 100. We published a report on this class as a paper "Piloting Data Engineering at Berkeley" at the Data Ed Workshop at *SIGMOD'22: SIGMOD International Conference on Data Management*, USA, 2022.
- ◇ **Co-Instructor for CS186**: Database Systems (Undergraduate Class), Fall 2020, Berkeley
  Students Enrolled: 620 (2020); co-taught with Alvin Cheung
- ◇ **Instructor for INFO290**: Human-in-the-loop Data Management aka Data Engineering (Graduate Lecture-based Class), Spring 2020, Berkeley
  Students Enrolled: 34 (2020)
- ◇ **Instructor for CS598**: Human-in-the-loop Data Management (Grad Seminar), Fall 2014, 2015, 2017, Illinois
  Students Enrolled: 16 (2014), 18 (2015), 18 (2017)
  List of Instructors Rated as "Excellent" for Fall 2015, Fall 2017
- ◇ **Instructor for CS411**: Introduction to Database Systems (Undergraduate Class), Spring 2016, 2017, 2018 Illinois
  Students Enrolled: 230 (2016), 230 (2017), 230 (2018)
- ◇ **Instructor for CS511**: Advanced Data Management (Graduate Lecture-based Class), Spring 2015, Illinois
  Students Enrolled: 45

Invited Talks

- ◇ Gong Show Talk at NorCal DB Day, October 2026
- ◇ Snowflake Sponsor Talk, "Reimagining Databases for Agents" at VLDB, September 2026
- ◇ Agentic Data Systems Workshop Keynote at VLDB, August 2026
- ◇ DASHSys Workshop Keynote at VLDB (declined), August 2026
- ◇ SynthAI Workshop Keynote at SIGMOD, June 2026
- ◇ HILDA Workshop Keynote at SIGMOD, June 2026
- ◇ Beyond SQL Workshop Keynote at ICDE, May 2026
- ◇ LLM+Vector Data Workshop Keynote at ICDE, May 2026
- ◇ IDSR Workshop Keynote at ICDE, May 2026
- ◇ AIDB Keynote at VLDB, September 2025
- ◇ NOVAS Workshop Keynote at SIGMOD, July 2025
- ◇ Stanford University, October 2024
- ◇ University of Illinois at Chicago, October 2024

- ◇ Disseminate Podcast, October 2024
- ◇ UC Irvine, September 2024
- ◇ New York University, February 2024
- ◇ Cornell University, November 2023
- ◇ LLM Panel at the Database Meeting, October 2023
- ◇ DEEM Workshop Keynote at SIGMOD, June 2023
- ◇ AI and Databases Panel at NorCal DB day, April 2023
- ◇ The Data Stack Show Podcast, April 2023
- ◇ Data Dividend, organized by IBM and the Economist, San Jose, February 2023
- ◇ SIAM Data Mining (SDM) Conference Keynote: April 2022
- ◇ VLDB New Researcher Symposium Keynote: August 2020
- ◇ Software Engineering Daily Podcast: May 2020
- ◇ IBM Distinguished Speaker Forum: May 2020 (Canceled due to COVID-19)
- ◇ DataCouncil SF Data Engineering Talk: March 2020 (Canceled due to COVID-19)
- ◇ Berkeley Information Access Seminar: November 2019
- ◇ Berkeley Institute of Design Talk: October 2019
- ◇ VLDB Early Career Research Contributions Award Talk: September 2019
- ◇ VLDB Summer School in China: July 2019
- ◇ UC Berkeley I School Colloquium: October 2018
- ◇ U Washington I School Colloquium: October 2018
- ◇ U Wisconsin CS Colloquium: March 2018
- ◇ UCSD CSE Colloquium: March 2018
- ◇ Princeton CS Colloquium: March 2018
- ◇ U Washington CS Colloquium: February 2018
- ◇ UC Berkeley EECS Colloquium: February 2018
- ◇ U Chicago CS Colloquium: February 2018
- ◇ U Maryland CS Colloquium: January 2018
- ◇ U Michigan CS Colloquium: January 2018
- ◇ U Penn Distinguished Lecture: November 2017
- ◇ U Waterloo Distinguished Lecture: October 2017
- ◇ TCDE Rising Star Award Talk at ICDE: March 2017
- ◇ IBM TJ Watson Invited Lecture: December 2016
- ◇ Northwestern EECS Distinguished Lecture: November 2016
- ◇ Midwest BigData Workshop: October 2016
- ◇ U. Illinois BigData Day: October 2016
- ◇ BIRS-CMO Workshop on Theory of Crowds and Networks: August 2016
- ◇ SIGKDD EI (Enterprise Intelligence) Workshop Keynote: August 2016
- ◇ National Center for Supercomputing Applications: August 2016
- ◇ University of Texas-Austin: November 2015
- ◇ University of Michigan: October 2015
- ◇ NIH Center Meeting: August 2015
- ◇ Google Mountain View Invited Talk: September 2014
- ◇ SIGKDD Dissertation Award Talk: August 2014
- ◇ SIGKDD IDEA (Interactive Data Exploration and Analysis) Workshop Invited Keynote: August 2014
- ◇ SIGMOD Jim Gray Award Keynote: June 2014
- ◇ UMass Amherst Invited Talk: April 2014
- ◇ MIT Data Analytics Workshop: April 2014
- ◇ INFORMS Conference Invited Talk: October 2013
- ◇ LinkedIn Data Science: August 2013
- ◇ UMichigan CS Colloquium: April 2013
- ◇ USC CS Colloquium: April 2013
- ◇ CMU SCS Colloquium: March 2013

- ◇ Illinois CS Colloquium: March 2013
- ◇ UC San Diego CS Colloquium: March 2013
- ◇ Princeton CS Colloquium: March 2013
- ◇ Harvard CS Colloquium: March 2013
- ◇ Georgia Tech CS Colloquium: March 2013
- ◇ U Maryland CS Colloquium: February 2013
- ◇ U Chicago CS Colloquium: February 2013
- ◇ Columbia University CS Colloquium: February 2013
- ◇ UC Santa Cruz Database Group Invited Talk: February 2013
- ◇ Duke University Invited Talk: January 2013
- ◇ Crowd-Crowd Workshop, Stanford: January 2013
- ◇ Amazon Machine Learning Group Invited Talk, Bangalore, India: September 2012
- ◇ IBM Research Invited Talk (HCI Group), Almaden, USA: July 2012
- ◇ MIT EECS Colloquium: April 2012
- ◇ Univ. of Washington CS Colloquium: February 2012
- ◇ IBM Research Invited Talk, Almaden, USA: March 2011
- ◇ Crowd-Crowd Workshop, UC Berkeley: June 2011
- ◇ Microsoft Research DMX Group, Redmond, USA: August 2010
- ◇ Yahoo! Research Invited Talk, Bangalore, India: September 2009
- ◇ UC Berkeley, Database Group Lunch Talk: March 2009
- ◇ Kosmix Corporation, Mountain View, USA: August 2008