Aditya G. Parameswaran

212 South Hall, Berkeley CA 94704
University of California, Berkeley

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

Employment

- **University of California, Berkeley**, Berkeley, CA.
  Associate Professor, School of Information. 2021—
  Associate Professor, Electrical Engineering and Computer Sciences. 2021—
  Faculty Affiliate, Berkeley Institute of Data Science (BIDS). 2022—
  Assistant Professor, School of Information. 2019—2021
  Assistant Professor, Electrical Engineering and Computer Sciences. 2019—2021
- **Ponder, Inc.**, Berkeley, CA.
  Co-founder and President. 2021—
- **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

Honors and Awards

- **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
- Runner Up, **Facebook Probability and Programming Research Award**, 2019
- **Best Demo Award**, ICDE for my paper "Faster, Higher, Stronger: Redesigning Spreadsheets for Scale". Awarded to two out of 24 demonstrations, 2019
- Army Research Office **Young Investigator Program (ARO YIP) Award**, 2018
- 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
- 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
- Listed in **"Instructors rated as Excellent by students @ Illinois"**, 2017, 2015

- Google **Faculty Research Award**, 2015, and **Focused Research Award**, 2017
- Selections for **best-papers-of-conference** for the following conferences (awarded to ~5 papers per year):
  - **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"
- Awards for my Ph.D. dissertation "Human-Powered Data Management", including:
  - **ACM SIGMOD** Jim Gray Dissertation Award, 2014
  - **ACM SIGKDD** Dissertation Award Runner Up, 2014
  - **Stanford's** Arthur Samuels Best Dissertation Award in Computer Science, 2014
- Best Demo 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"

Broad Interests Interactive Data Exploration, Human-Centered Data Science, Data Management, Databases

Relevant Expertise My work centers on the design of *human-centered and scalable data science tools* by synthesizing techniques from multiple fields: **databases, data mining,** and **human computer-interaction**. I've published over **100 papers** in the top-tier venues of these fields with an **h-index of 40+**, an **i-10 index of 75+**, and nearly **5000** citations. My research begins with a thorough exploration of the foundational principles, followed by the design of practical, scalable, and usable systems and algorithms.

Research Themes
- **Visual Analytics [VA]:** Scalable visualization generation and recommendation on large datasets
- **Interactive Analytics [IA]:** Interactively analyzing large volumes of data
- **Optimized Human Computation [OC]:** Optimizing crowdsourcing for large-scale data processing
- **Simplified Information Extraction [IE]:** Reducing human involvement in web information extraction
- **Enhanced Recommendation Systems [RS]:** Using context to improve recommender systems
- **Miscellaneous [MS]**

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

1. [IA, VA] 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.
2. [IA] 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.
3. [IA, VA] 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.
4. [IA] 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.
5. [IA, VA] 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.
6. [IA] 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

7. [IA, VA] 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
8. [IA] 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.
9. [IA, VA] 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**.
10. [IA] 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
11. [IA, VA] 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.
12. [IA] 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.
13. [IA] 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.
14. [IA] 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.
15. [VA] 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)**.
16. [IA] 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.
17. [IA] 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.
18. [IA] 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.
19. [OC] 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.
20. [IA, VA] **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**.
21. [VA] 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.
22. [IA] 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.
23. [IA, VA] 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.
24. [IA, VA] C. Yen, **A. Parameswaran**, W. Fu. An Exploratory User Study of Visual Causality Analysis. *EuroVis'19: 21st Eurographics Conference on Visualization*, Porto, Portugal, 2019.
25. [VA, IA] 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.

26. [IA] 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.
27. [IA] 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.
28. [IA, VA] 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.
29. [IA, IE] 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%.
30. [IA] 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.
31. [IA] 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.
32. [IA] 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.
33. [MS] 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.
34. [IA] 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.
35. [IA] 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%.
36. [IA] 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**.
37. [OC] 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%.
38. [VA] 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%.
39. [IE] 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%.
40. [MS] 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**.
41. [VA] 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. [IE] 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%.
43. [IA] 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%.

44. [VA] 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%.
45. [IA] 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%.
46. [OC] 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%.
47. [IA] M. Joglekar*, H. Garcia-Molina, and **A. Parameswaran**. Interactive Data Exploration with Smart Drill-down. *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**.
48. [VA] M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *SIGMOD Record*, December 2016.
49. [OC] **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
50. [OC] 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
51. [OC] 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.
52. [IA] 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%.
53. [VA] 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%.
54. [OC] 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%.
55. [OC] 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%.
56. [IA] 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.
57. [IA] 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%.
58. [OC] 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%.
59. [IA] 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.
60. [OC] **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%.
61. [OC] 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**.

62. [OC] **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%.
63. [IE] **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%.
64. [OC] 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%.
65. [OC,IE] 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.
66. [OC, IE] 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.
67. [OC] **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%.
68. [OC,IE] 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**.
69. [OC] 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%.
70. [OC] **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%.
71. [MS] 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%.
72. [OC] 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.
73. [IE] **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%.
74. [OC, IE] **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%.
75. [OC] **A. Parameswaran** and N. Polyzotis. Answering Queries using Databases, Humans, and Algorithms, *CIDR '11: Conf. on Innovative Data Management (CIDR)*, Asilomar, USA, 2011.
76. [RS] G. Koutrika, H. Garcia-Molina, and **A. Parameswaran**. Information Seeking: Convergence of Search, Recommendations, and Advertising, *Communications of the ACM (CACM)*, November 2011.
77. [RS] **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.
78. [RS] **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%.
79. [IE] **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**.
80. [RS] **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%.
81. [IE] 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%.

82. [RS] 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.
83. [MS] 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**

84. [OC] 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**

85. [VA] 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.
86. [IA, VA] 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)**.
87. [OC] 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.
88. [IA] 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.
89. [VA] 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.
90. [IA] 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.
91. [OC] Doris Lee*, Akash Das Sarma*, **A. Parameswaran**. Aggregating Crowdsourced Image Segmentations (Short Paper). *HCOMP '18: 6th Int'l Conf. on Human Computation and Crowdsourcing*), Zurich, Switzerland, 2018.
92. [IA] 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.
93. [IA] 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**.
94. [IA] 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.
95. [IA] 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.
96. [IA] 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.

97. [MS] S. Koltani, S. Wang, **A. Parameswaran**. GeoHashViz: Interactive Analytics for Mapping Spatiotemporal Diffusion of Twitter Hashtags (Poster). *XSEDE '15*, USA, 2015.
98. [OC] 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.
99. [VA] M. Vartak*, S. Huang*, T. Siddiqui*, S. Madden, and **A. Parameswaran**. Towards Visualization Recommendation Systems, *DSIA '15: Workshop on Interactive Analysis*, Chicago, 2015.
100. [VA] **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.
101. [VA] 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.
102. [OC] **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.
103. [OC] 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.
104. [OC] 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.
105. [RS] **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%.
106. [MS] 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.

◇ **Preprints**

107. [IA] 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
108. [IA] T. Wattanawaroon*, S. Macke*, **A. Parameswaran**. Towards a Theory of Data-Diff: Optimal Synthesis of Succinct Data Modification Scripts. Technical Report. December 2017
109. [OC] 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
110. [IA] L. Battle, T. Benson, **A. Parameswaran**, and E. Wu. Indexing Cost-Sensitive Prediction. Technical Report. August 2014

◇ **Other Writings**

111. [MS] **A. Parameswaran**. Visual Data Exploration: A Fertile Ground for Data Management Research. ***SIGMOD Blog***, June 2018.
112. [MS] D. Lee*, T. Siddiqui*, K. Karahalios, **A. Parameswaran**. Effortless Exploration with Zenvisage 0.2, *Medium*, November 2017.
113. [MS] **A. Parameswaran**. Enabling Data Science for the Majority, ***O'Reilly Blog***, October 2017.
114. [MS] A. Das Sarma*, **A. Parameswaran**. Crowdsourcing in Practice: Our Findings, *Medium*, September 2017.
115. [MS] S. Huang*, L. Xu*, A. Elmore, **A. Parameswaran**. Painless Data Versioning for Collaborative Data Science, *Medium*, September 2017.
116. [MS] S. Rahman* and **A. Parameswaran**. Draw Conclusions Early with Incvisage, *Medium*, September 2017.
117. [MS] T. Siddiqui* and **A. Parameswaran**. Towards Automating Insight, *Medium*, September 2017.
118. [MS] **A. Parameswaran**. Three Tools for Human-in-the-loop Data Analytics, *Medium*, January 2017.

119. [MS] M. Vartak* and **A. Parameswaran**. Instantly Visualizing What's Different, *Medium*, September 2016.
120. [MS] Y. Gao* and **A. Parameswaran**. On Optimally Squishing Large Datasets, *Medium*, August 2016.
121. [MS] A. Das Sarma* and **A. Parameswaran**. Revisiting Quality Management for Crowdsourcing: A Fresh Take, *Medium*, July 2016.
122. [MS] **A. Parameswaran**. Data Scavengers, not Research Parasites, *Medium*, January 2016.
123. [MS] A. Marcus, **A. Parameswaran**. The Fourth Industrial Revolution, *Medium*, January 2016.

Major Software Releases

- **MLTrace**: A bolt-on ML pipeline monitoring and observability solution. **8K Downloads, 360 GitHub stars** as of March 2022. 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. **1.5M Downloads, 6.9K GitHub Stars** as of March 2022. 2020–. `https://github.com/modin-project/modin`
- **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. **110K Downloads, 3.5K GitHub Stars** as of March 2022. 2020–. `https://github.com/lux-org/lux`
- **Nbsafety**: 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. **115K Downloads, 100 GitHub Stars** as of March 2022. 2020–. `https://github.com/nbsafety-project/nbsafety`
- **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`
- **Squish**: A near-optimal structured data compression system. Squish identifies correlations between attributes to compress relational datasets both vertically as well as horizontally. 2016–2017. `github.com/Preparation-Publication-BD2K/db_compress`
- **Populace**: Software releases of optimized implementations of various crowdsourced data processing algorithms and systems. 2015–2017. `http://populace-org.github.io`

Funding

I have been part of teams that have brought in a total of **over 15M in grant funding** *into my institution; and* **over 5.75M as PI or Co-PI** *from a variety of federal and industry funding sources.*

- Co-director, **EPIC** (Effective Programming, Interaction, and Computation with) **Data Lab**, supported via industry sponsors (details TBD). 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 [All students at Berkeley unless indicated otherwise.]
  - · Doris Lee,
    Title: "*Designing Automated Assistants for Visual Data Exploration*"
    2017—2021 (Winner: **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.
  - · 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, Linea Labs.
  - · Devin Petersohn,
    Title: "*Dataframe Systems: Theory, Architecture, and Implementation*"
    2015—2021 (Winner: **NSF GRFP**, **James Tullock Memorial Scholarship Award**)
    First employment: CTO, Ponder.

- 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.
- 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: TBD.
- 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: Researcher at Megagon Labs.
- Illinois: Liqi Xu,
  Title: "*New Capabilities for Large-scale Exploratory Data Analysis*"
  2015–2020
  First employment: Research Scientist at Facebook.
- 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.
- 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.

◇ PhD Students (Current Advising or Co-advising)

- Shreya Shankar, 2021— (Winner: **Hertz Foundation Fellowship Finalist**);
- Nithin Chalapathi, 2021— (Coadvisor: Alvin Cheung);

◇ Postdocs (Current or Past)

- Dixin Tang, 2021—;

◇ MS Thesis Students

- Chris De Leon, 2021–;
- Priyam Mohanty, 2021–;
- Kunal Aggarwal, 2021–;

- 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);

◇ PhD/MS Students (Informal or Past advising)

- Jennifer Momoh, 2020–21; Thyne Boonmark, 2019–21; 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); Himel Dev, 2015—16;
- MIT: Albert Kim, 2013—16; Manasi Vartak, 2013—15;
- 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: Micah Yong, 2020–; Jonathan Yun, 2021–; Connor Lien, 2022–; Joshua Wu, 2021–; Lawrence Chen, 2022–; Pragya Kallanagoudar, 2022–; Todd Yu, 2021–; Christina Fan, 2022–; Eve Yi, 2022–; Chris Seo, 2022–;
- Past students opting for Ph.D. studies:
  - 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 industry employment: 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)

External Service Activities

◇ **Steering Committees:**

- Human-in-the-Loop Data Analytics (HILDA) Workshop Steering Committee at the ACM SIGMOD Int'l Conference on Management of Data, 2016—
- Data Systems for Interactive Analysis (DSIA) Workshop Steering Committee at the IEEE VIS Int'l Conference on Visualization, 2018—

◇ **Conference Chair Positions:**

- Chair, Workshops, SIGMOD 2021.
- Sponsor Chair (US), VLDB 2021.
- Chair, Workshops, SIGMOD 2020.
- Area Chair, SIGMOD 2020.
- Associate Chair, HCOMP 2019.
- 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

◇ **Editorships:**

- 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.

- Guest Editor, ACM XRDS Magazine (Issue on "Big Data"), with 15 invited articles from experts, 2012

◇ **Tutorials:**

- 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.
- Invited full day Tutorial at the VLDB Summer School at Renmin University, Beijing, 2019.

◇ **Working Groups:**

- NIH BD2K Commons Working Group, 2015–18.
- NIH BD2K Machine Learning Working Group, 2016–18.

◇ **Award Committees:**

- SIGMOD Jim Gray Dissertation Award Committee, 2015
- SIGMOD Best Demo Award Committee, 2014
- PURE Undergraduate Research Award Committee, 2014

◇ **Panel Participation:**

- 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

◇ **Review Panels:**

- 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:**

- VLDB Demo 2019
- ICDE Demo 2018
- SIGMOD Demo 2017
- 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:** BIRS-CMO Workshop on "Theory and Models of Crowds and Networks", Oaxaca, Mexico, 2016

Internal Service Activities

◇ Berkeley, I School Faculty Equity Advisor 2021
◇ Berkeley, I School PhD Admissions 2021
◇ Berkeley, I School MIMS Admissions 2021

- Berkeley, CDSS DataHub Space Committee 2020–21
- Berkeley, EECS Graduate Admissions Committee 2019–21
- Berkeley, I School MIDS Curriculum Committee 2019–20
- Berkeley, 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 CS194/290**: Data Engineering (Undergraduate/Graduate Class), Spring 2021, Berkeley
  Students Enrolled: 125 (2021)
- **Instructor for CS186**: Database Systems (Undergraduate Class), Fall 2020, Berkeley
  Students Enrolled: 620 (2020)
- **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

- 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