Eugene Wu **ewu@cs.columbia.edu**
500 West 120 St, Room 421 NY, NY 10027

RESEARCH INTERESTS

I am broadly interested in the design of data management systems, and the application of database concepts, to greatly extend the ability for domain-experts and normal users to work with data. This draws upon the areas of database optimization, human-data-interaction, data provenance, visualization, and interface design.
Prepared April 6, 2021

EDUCATION

Winter 2014 **Massachusetts Institute of Technology**, Cambridge, MA
Ph.D., Electrical Engineering and Computer Science
Advisor: Samuel Madden
Dissertation: Explaining Data in Visual Analytic Systems

May 2010 **Massachusetts Institute of Technology**, Cambridge, MA
M.S., Electrical Engineering and Computer Science
Advisor: Samuel Madden
Dissertation: Shinobi: Insert-aware Partitioning and Indexing Techniques For Skewed Database Workloads

Spring 2007 **UC Berkeley**, Berkeley, CA
B.S., Electrical Engineering and Computer Science

PROFESSIONAL EXPERIENCE

2021 Spring On parental leave

2020–2021 **Columbia University**, NY, NY
Associate Professor (untenured) – Computer Science
Co-director – Center for Data, Media & Society, Columbia Data Science Institute
Co-advisor – Columbia Journalism & Computer Science Dual Degree Program

2015–2020 **Columbia University**, NY, NY
Assistant Professor – Computer Science
Co-director – Center for Data, Media & Society, Columbia Data Science Institute
Co-advisor – Columbia Journalism & Computer Science Dual Degree Program

2015 **UC Berkeley**, Berkeley, CA
Visiting Scholar – AMPLab

2008–2014 **Massachusetts Institute of Technology**, Cambridge, MA
Ph.D. Student – CSAIL

2007-2008 **Google Inc.**, Mountain View, CA
Research Intern – Data Management Group

AWARDS

2021 **Google Research Scholar Award**

Amazon Faculty Award

2020 **Very Large Data Bases (VLDB) Distinguished Reviewers Award**

2019 **Conference on Innovative Data Systems Research (CIDR) Best Gong-show Award** for best 5-minute short talk award.

SIG Management of Data (SIGMOD) Exceptional Service Award

NSF CAREER Award

2018 **Very Large Data Bases (VLDB) 10 Year Test-of-Time Award** for the WebTable project, which analyzed over 14 billion HTML tables to collect the largest corpus of database-like tables on the web.

Google Faculty Award for development of deep neural inspection, to identify high-level logic in neural network models.

Amazon Faculty Award for development of human-in-the-loop techniques to understand machine learning models.

2016 **SIG Management of Data (SIGMOD) best demo award** for demonstration of model-aware database cleaning techniques.

TEACHING EXPERIENCE

At Columbia University

2020 Fall — *Instructor, Topics in Database Research (W6113, 15 students)*
http://w6113.github.io

2020 Spring — *Instructor, Systems for Human Data interaction (W6998, 13 students)*
https://columbiaviz.github.io/2020s_w6998/

2019 Spring — *Instructor, Introduction to Databases (W4111, 73 students)*
http://w4111.github.io

2019 Spring — *Instructor, Database Research Topics (W6113, 13 students)*
http://w6113.github.io

2018 Fall — *Instructor, Introduction to Databases (W4111, 107 students)*
http://w4111.github.io

2018 Spring — *Instructor, Database Topics in Research & Practice (W6998.005, 9 students)*
http://columbiadb.github.io

2018 Spring — *Co-Instructor, Computing Systems for Data Science (W4121, 158 students)*
http://w4121.github.io

2017 Spring — *Instructor, Interactive Data Exploration Systems (W6998.002, 10 students)*
http://columbiaviz.github.io

2017 Spring — *Co-Instructor, Computing Systems for Data Science (W4121, 124 students)*
http://w4121.github.io

2016 Fall — *Instructor, Introduction to Databases (W4111, 139 students)*
http://w4111.github.io

2016 Fall — *Instructor, Introduction to Databases CVN sec. (W4111, 39 students)*
http://w4111.github.io

2016 Spring — *Co-Instructor, Big Data Systems (W4121, 79 students)*
http://w4121.github.io

2015 Fall — *Instructor, Introduction to Databases (W4111, 108 students)*
http://w4111.github.io

Prior to Columbia University

2013 Fall — *Instructor, From Ascii To Answers (MIT 6.885)*
I co-developed and instructed MIT's first Big Data course focused on large scale data analysis tools and techniques. Topics ranged from data cleaning and integration, large-scale systems like Hadoop, to scalable visualization techniques. We developed eight labs to give students hands-on experience with the systems covered in class. The course is freely available online at http://github.com/mitdbg/asciiclass

2012 Spring — *Instructor, Introduction to Data Analysis*
I co-developed and taught an Introduction to Data Analysis course to approximately 20 students during MIT's Independent Activities Period in January. The course is freely available online at http://dataiap.github.io

2011 – 2012 — *Head of Curriculum, MEET*
MEET is a 3-year technology program and peace initiative that teaches Israeli and Palestinian high school students. I organized curriculum preparation for each year's incoming instructors. I also successfully migrated the organization from a Java-based curriculum to a Python-oriented one and developed the lesson plans for the transition.

2010 Fall — *Teaching Assistant, Database Systems (MIT 6.830)*
I assisted in writing and grading the assignments and projects.

2010 Summer — *Instructor, MEET*
I mentored a group of 30 Israeli and Palestinian high school students as part of the MIT MEET program, a peace initiative in the Middle East centered around teaching computer science.

2010, 2011 Spring — *Instructor, Introduction to Java Course (MIT 6.S092)*
I instructed a class of 50 students in an introduction to the Java programming language. MIT does not have such an introductory course, so this course is taken by many MIT undergraduates to prepare them for 6.004, a core course that assumes proficiency in Java. The course is freely available online at http://bit.ly/alvK9m

2006 Fall — *Teaching Assistant, Database Systems (UCB CS186)*
I taught approximately 30 students in weekly discussion sections. I assisted in writing and grading the assignments and projects.

COMMUNITY SERVICE

2022	VLDB Associate Editor
2021	On parental leave
	VLDB Associate Editor
2020	VLDB PC
	Department of Energy Early Career PC
2019	NY DB Day Workshop Co-Chair
	SIGMOD PC
	SIGMOD Student Research Competition Co-chair
2018	ICDE PC
	SIGMOD PC
	HILDA Co-chair
	SIGMOD New Researcher Symposium Co-chair
	SIGMOD Publicity Co-chair
	Dagstuhl 17461: Connecting Visualization and Data Management Research
2017	ICDE Area Chair
	WWW PC
	SIGMOD Demo PC
	SIGMOD PC
	VLDB PC
	HILDA PC
	SSDBM PC
	HCOMP PC
	CHI Reviewer
2016	InfoVis Reviewer
	HILDA PC
	CHI Reviewer
	InfoVIS Reviewer
	NEDBDay Co-Chair
	SIGMOD travel award Co-chair
2015	SIGMOD travel award Co-chair
2014	DATA4U PC

Glossary

SIGMOD	ACM SIGMOD/PODS International Conference on Management of Data
VLDB	International Conference on Very Large Data Bases
ICDE	IEEE Conference on Data Engineering
CIDR	Conference on Innovative Data Systems Research
CHI	ACM CHI Conference on Human Factors in Computing Systems
WWW	The Web Conference
SSDBM	International Conference on Scientific and Statistical Database Management
HCOMP	AAAI Conference on Human Computation and Crowdsourcing
HILDA	Human-in-the-Loop Workshop at SIGMOD
InfoVIS	IEEE Information Visualization conference
NEDBDay	New England Database Day
NYDBDay	New York Database Day
DATA4U	Workshop on Bringing the Value of “Big Data” to Users

COLUMBIA SERVICE

2021 On parental leave
Data Science Institute Co-Chair Center for Data, Media & Society

2020 Data Science Institute Co-Chair Center for Data, Media & Society
CS Chair for CS-Journalism Dual Degree
Faculty mentor for Columbia Virtual Campus
DSI Data Art Contest Co-organizer
Computer Science Graduate Admissions Committee

2019 Data Science Institute Co-Chair Center for Data, Media & Society
CS Chair for CS-Journalism Dual Degree
DSI Data Art Contest Co-organizer
Computer Science Faculty Recruiting Committee
Computer Science Graduate Admissions Committee

2018 Data Science Institute Co-Chair Center for Data, Media & Society
Data Science Institute Masters Curriculum
Data Science Institute Center Committee: Center for Computing Systems for Data-Driven Science
CS Chair for CS-Journalism Dual Degree
Computer Science Graduate Admissions Committee

2017 Data Science Institute Co-Chair Center for Data, Media & Society
Data Science Institute Center Committee: Center for Computing Systems for Data-Driven Science
CS Chair for CS-Journalism Dual Degree
Computer Science Graduate Admissions Committee

2016 Data Science Institute Co-Chair Center for New Media
CS Chair for CS-Journalism Dual Degree
Computer Science Graduate Admissions Committee

2015 Computer Science Graduate Admissions Committee

INVITED TALKS

2021 Systems for Human Data Interaction *Keynote at SEAS DATA Workshop at VLDB*
Systems for Human Data Interaction *CMU, Vaccination DB Talk Series*

2020 Systems for Human Data Interaction *Tufts University*
Systems for Human Data Interaction *University of Chicago*
Systems for Human Data Interaction *UCSD*
Systems for Human Data Interaction *UC Berkeley*

2019 Systems for Human Data Interaction *Tsinghua University, Beijing, China*
Systems for Human Data Interaction *China Big Data Workshop, Beijing, China*
Tutorial: Towards Democratizing Relational Data Visualization *VLDB Summer School, Renmin University, Beijing, China*
Precision Interfaces *SIGMOD in Amsterdam, Netherlands*
Tutorial: Towards Democratizing Relational Data Visualization *SIGMOD in Amsterdam, Netherlands*
Crazier Innovations in Databases ⋈ Reinforcement-learning Research *CIDR in Monterey, CA*
Human Data Interaction in the WuLab *UCLA*

2018 Human Data Interaction in the WuLab *Columbia Lamont Earth Institute*
Human in the loop data analysis panel *NYDBDay, NY*
Closing the loop on data analysis *NYDBDay, NY*
Closing the loop on data analysis *UMass Amherst*
Closing the loop on data analysis *BlueCore, NY*
At a Glance: Approximate Entropy as a Measure of Line Chart Visualization Complexity *InfoVIS, Berlin, Germany*
Closing the loop on data analysis *Hasso Platner Institute, Potsdam, Germany*
Closing the loop on data analysis *Data Science Institute, Columbia University*
Closing the loop on data analysis *Harvard University*
Closing the loop on data analysis *MIT*
Closing the loop on data analysis *Google Research NYC*

2017 Closing the loop on data analysis *AT&T Research NYC*
Closing the loop on data analysis *University of Fribourg, Switzerland*
Closing the loop on data analysis *Dagstuhl, Germany*
Closing the loop on data analysis *UIUC*
Closing the loop on data analysis *U. Chicago*
Closing the loop on data analysis *U of Wisconsin-Madison*
Databases and Data Visualization *2Sigma, NY*
ICDE Panel: The Case for Small Data *ICDE in San Diego, CA*
CIDR: Chat-oriented Innovations in Database Research *CIDR in Santa Cruz, CA*
Combining Design and Performance in a Data Visualization Management System *CIDR in Santa Cruz, CA*

2016 Closing the loop on data analysis *Foundations of Data Science, DSI, Columbia University*
Closing the loop on data analysis *Sense, Collect & Move, DSI, Columbia University*
Closing the loop on data analysis *Brown University*
Towards Perception-aware Interactive Data Visualization Systems *DSIA in Chicago, IL*
Closing the loop on data analysis *Egelston Seminar at Columbia University*
Closing the loop on data analysis *IGERT Seminar at Columbia University*
PFunk-H: Approximate Query Processing using Perceptual Models *HILDA at SIGMOD, SF, CA*
Closing the loop on data analysis *IBM in Yorktown, NY*
Provenance in Big Data *MIT Big Data Workshop, Boston, MA*
Closing the loop on data analysis *Stats at Columbia University*
Closing the loop on data analysis *Rutgers University*

2015 Closing the loop on data analysis *IGERT Seminar at Columbia University*
VLDB Panel: Designing for Interaction: Broadening our View of Working with Data *VLDB in Hawaii*

STUDENTS

Postdocs
Thibault Sellam. Now at Google.

Ph.Ds
Fotis Psallidas (Advisor). Now at Microsoft Research Gray Systems Lab
Thesis: Physical Plan Instrumentation in Databases: Design Principles and Applications
Yiru Chen (Advisor)
Lampros Flokas (Advisor)
Zachary Huang (Advisor)
Haneen Mohammed (Advisor)
Lana Ramjit, UCLA (Co-advised with Ravi Netravali at UCLA).
Yiliang Shi (Co-advised with Carl Vondrick). Graduated with M.S. degree.

Orestis Polychroniou (Thesis Committee).
Thesis: Analytical Query Execution Optimized for all Layers of Modern Hardware
Ioannis Paparrizos (Thesis Committee).
Thesis: Fast, Scalable, and Accurate Algorithms for Time-Series Analysis
Daniel Miau (Thesis Committee).
Thesis: Personalized Navigational Instruments for Map User Interfaces
Wangda Zhang (Thesis Committee).
Thesis: Optimizing Query Processing Under Skew

Yifan Wu, UC Berkeley (Thesis Committee). Primary advisor: Joe Hellerstein
Mohammad Mahdavi, TU Berlin (Thesis Committee). Primary advisor: Ziawasch Abedjan
Marianne Procopio, Tufts (Thesis Committee). Primary advisor: Remco Chang

MS
Daniel Alabi (Advisor). Now Ph.D. at Harvard University
Zhengjie Miao (Advisor). Now Ph.D. at Duke University
Gabriel Ryan (Advisor). Now Ph.D. at Columbia University
Haoci Zhang (Advisor). Now at Facebook
Tejas Dharamsi (Advisor). Now at Trifacta
Sharan Suryanarayanan (Advisor). Now at Facebook
Charlie Summers (Advisor).
Lucy X Wang (Thesis Committee).
Thesis: Modeling and Predicting the Dynamics of Clicks from Social Media

Undergrads
Hamed Nilforoshan (Advisor). Ph.D. at Stanford.
Ian Yiran Huang (Advisor). Ph.D. at Stanford.
Robert Netzorg (Advisor). Ph.D. at UC Berkeley.
Kevin Lin (Advisor). AI2, Ph.D. at UC Berkeley.
Lauren Arnett (Advisor)
Jeffrey Huang (Advisor)
Gitika Bose (Advisor)
Tejit Pabari (Advisor)
Amita Shukla (Advisor)
Condor Shou (Advisor)
Viraj Rai (Advisor)
Maneet Khaira (Advisor)
Sagar Lal (Advisor). M.S. at Columbia.
Rodolfo Raimundo (Advisor)
Jacob Fisher (Advisor)
Rahul Khanna (Advisor). M.S. at USC.

High School	Alex Studer
	Joey O Connor

GRANTS

Total Awarded $1,965,038

Awarded

2021 Google Research Scholar Award
PI: Eugene Wu.

Amazon Faculty Award
PI: Eugene Wu.

2020 NSF REU: CAREER: Visual Database Interfaces
PI: Eugene Wu

III: Small: Bringing database query optimization to data intensive applications
PI: Ken Ross. Co-PI: Eugene Wu

Columbia SIRS: Filling "Data Voids" to Protect the Food Supply Against Climate Change
PI: Eugene Wu. Co-PI: Lydia Chilton. Co-PI: Daniel Osgood, Columbia University Earth Institue

2019 NSF CAREER: Visual Database Interfaces
PI: Eugene Wu

2018 Google Faculty Award
PI: Eugene Wu

Amazon Faculty Award
PI: Eugene Wu

Columbia DSI Scholars: Data Cleaning, Data Reconciliation and Data Visualization
PI: Eugene Wu. Co-PI: Daniel Osgood, Columbia University Earth Institute

2017 Columbia Alliance: Perceptual Functions for Faster Interactive Visualizations
PI: Eugene Wu. Co-PI: Yanlei Diao, Ecole-Polytechnique

NSF: I-Corps: Internet of Things Monitoring System
PI: Eugene Wu.

2016 NSF: ACM SIGMOD Conference 2016: Student Activities and Travel Support
PI: Eugene Wu.

NSF IIS: Medium: Collaborative Research: Composing Interactive Data Visualizations
PI: Joseph Hellerstein, Berkeley. Co-PI: Jeff Heer, University of Washington. Co-PI: Eugene Wu.

2015 III: Small: Collaborative Research: Towards Interactive Data Visualization Management Systems
PI: Arnab Nandi, Ohio State University. Co-PI: Eugene Wu.

NSF REU: Development of Graphical Perception as a Service

PI: Eugene Wu.

Glossary

NSF	National Science Foundation
RI	Robust Intelligence
III	Information Integration and Informatics
REU	Research Experiences for Undergraduates

PUBLICATIONS

H-index: 31.
Total citations: 5,486.
Citation counts are reported from http://scholar.google.com when exceeding ten citations. Acceptance rates are reported for refereed conference papers when available.

Authorship Conventions

Conferences are the primary publication venue in data management. Senior authors are typically listed last. Advisees are listed in **bold**. Eugene Wu is underlined.

In Progress or In Review

[1] Y. Wu, R. Chang, J. Hellerstein, A. Satyanarayan, E. Wu. "*DIEL: Interactive Visualization Beyond the Here and Now.*" VIS 2021 (in review).

[2] Y. Wu, Y. Liu, **L. Flokas**, J. Wang, E. Wu. "*Private Federated Explanation of Inference Queries.*" VLDB 2021 (in progress).

[3] B. Lockhard, J. Wang, E. Wu. "*Explaining SQL-ML Queries with Bayesian Optimization.*" VLDB 2021 (under revision).

[4] E. Wu. "*Visual Composition Algebra.*" TVCG 2021 (in progress).

[5] **J. Fisher**, R. Chang, E. Wu. "*Dynamic Breakpoints for Y-axis Scales.*" InfoVIS Workshop 2021 (in progress).

[6] **L. Ramjit**, S. Mitra, R. Netravali, E. Wu. "*Physical Visualization Design.*" VLDB 2021 (in progress).

[7] **Z. Huang**, E. Wu. "*Reptile: Aggregation-level Explanations for Hierarchical Data.*" VLDB 2021 (in review).

[8] **Y. Chen**, E. Wu. "*Interactive Visualization Generation From Queries.*" VLDB 2021 (in progress).

Full Publications

[9] **R. Netzorg**, **L. Arnett**, A. Chaintreau, E. Wu. "*PopFactor: Live-Streamer Behavior and Popularity.*" ICWSM 2021.

[10] M. Procopio, A. Mosca, C. Scheidegger, E. Wu, R. Chang. "*Impact of Cognitive Biases on Progressive Visualization.*" TVCG 2021. (acceptance rate: 24.6%)

[11] F. Neutatz, B. Chen, Z. Abedjan, E. Wu. "*From Cleaning Before ML to Cleaning For ML.*" Invited, IEEE Data Engineering Bulletin 2021.

[12] **H. Mohammed**, Z. Wei, R. Netravali, E. Wu. "*Continuous Prefetch for Interactive Data Applications.*" VLDB 2020. (acceptance rate: 24.8%)

[13] Y. Wu, **L. Flokas**, J. Wang, E. Wu. "*Complaint-driven Training Data Debugging for Query 2.0.*" SIGMOD 2020. (acceptance rate: 26.9%)

[14] **Y. Chen**, E. Wu. "*Monte Carlo Tree Search for Generating Interactive Data Analysis Interfaces.*" Intelligent Process Automation (IPA) 2020.

[15] **L. Ramjit**, M. Interlandi, E. Wu, R. Netravali. "*Acorn: Aggressive Result Caching in Spark SQL.*" SOCC 2019. (acceptance rate: 25%)

[16] N. Tang, E. Wu, G. Li. "*Towards Democratizing Relational Data Visualization.*" SIGMOD 2019 Tutorial.

[17] **Q. Zhang**, **H. Zhang**, **V. Rai**, **T. Sellam**, E. Wu. "*Precision Interfaces.*" SIGMOD 2019. (acceptance rate: 20%)

[18] P. Wang, J. Wang, R. Shea, E. Wu. "*Progressive Deep Web Crawling Through Keyword Queries For Data Enrichment.*" SIGMOD 2019. (acceptance rate: 20%)

[19] **T. Sellam**, **K. Lin**, **I. Y. Huang**, M. Yang, C. Vondrick, E. Wu. "*DeepBase: Deep Inspection of Neural Networks.*" SIGMOD 2019. (acceptance rate: 20%, citations: 11)

[20] M. Cafarella, A. Halevy, D. Z. Wang, H. Lee, J. Madhavan, C. Yu, E. Wu. "*Ten Years of Web Tables.*" PVLDB 2018 Invited Paper,. (citations: 32)

[21] **G. Ryan**, A. Mosca, R. Chang, E. Wu. "*At a Glance: Approximate Entropy as a Measure of Line Chart Visualization Complexity.*" InfoVIS 2018. (acceptance rate: 25.13%, citations: 13)

[22] **H. Nilforoshan**, E. Wu. "*Leveraging Quality Prediction Models for Automatic Writing Feedback.*" ICWSM 2018. (acceptance rate: 20%)

[23] **T. Sellam**, **K. Lin**, **I. Y. Huang**, C. Vondrick, E. Wu. "*"I Like the Way You Think!" Inspecting the Internal Logic of Recurrent Neural Networks.*" SysML 2018.

[24] **F. Psallidas**, E. Wu. "*Smoke: Fine-grained Lineage at Interactive Speeds.*" VLDB 2018. (acceptance rate: 21%, citations: 43)

[25] L. Sun, M. J. Franklin, J. Wang, E. Wu. "*Skipping-oriented Partitioning for Columnar Layouts.*" VLDB 2017. (acceptance rate: 18.6%, citations: 26)

[26] E. Wu, **F. Psallidas**, **Z. Miao**, **H. Zhang**, L. Rettig, Y. Wu, **T. Sellam**. "*Combining Design and Performance in a Data Visualization Management System.*" CIDR 2017. (citations: 26)

[27] X. Wang, A. Meliou, E. Wu. "*QFix: Diagnosing errors through query histories.*" SIGMOD 2017. (citations: 25)

[28] D. Haas, J. Wang, E. Wu, M. J. Franklin. "*CLAMShell: Speeding up Crowds for Low-latency Data Labeling.*" VLDB 2016. (acceptance rate: 35.5%, citations: 65)

[29] A. A. Bhattacharya, D. Hong, D. Culler, J. Ortiz, K. Whitehouse, E. Wu. "*Automated Metadata Construction to Support Portable Building Applications.*" BuildSys 2015. (citations: 64)

[30] E. Wu. "*Explaining Data in Visual Analytic Systems.*" Doctoral Thesis 2015.

[31] E. Wu, L. Battle, S. Madden. "*The Case for Data Visualization Management Systems.*" VLDB 2014. (citations: 85)

[32] A. Jindal, P. Rawlani, E. Wu, S. Madden, A. Deshpande, M. Stonebraker. "*Vertexica: Your Relational Friend for Graph Analytics!.*" SIGMOD 2014 demo. (citations: 63)

[33] A. Cheung, L. Ravindranath, E. Wu, S. Madden, H. Balakrishnan. "*Mobile applications need Targeted Micro-updates.*" APSYS 2013.

[34] E. Wu, S. Madden. "*Scorpion: Explaining Away Outliers in Aggregate Queries.*" VLDB 2013 (Best-of). (acceptance rate: 22.7%, citations: 211)

[35] E. Wu, S. Madden, M. Stonebraker. "*SubZero: a Fine-Grained Lineage System for Scientific Databases.*" ICDE 2013 (Best-of). (citations: 43)

[36] A. Marcus, E. Wu, D. Karger, S. Madden, R. Miller. "*Human-powered Sorts and Joins.*" VLDB 2012. (acceptance rate: 20.3%, citations: 334)

[37] E. Wu, S. Madden. "*Partitioning Techniques for Fine-Grained Indexing.*" ICDE 2011. (acceptance rate: 20%, citations: 39)

[38] A. Marcus, E. Wu, D. Karger, S. Madden, R. Miller. "*Demonstration of Qurk: A Query Processor for Human Operators.*" SIGMOD 2011. (citations: 51)

[39] E. Wu, C. Curino, S. Madden. "*No Bits Left Behind.*" CIDR 2011.

[40] A. Marcus, E. Wu, S. Madden, R. Miller. "*Crowdsourced Databases: Query Processing with People.*" CIDR 2011. (citations: 248)

[41] C. Curino, E. Jones, R. Popa, N. Malviya, E. Wu, S. Madden, H. Balakrishnan, N. Zeldovich. "*Relational Cloud: A Database-as-a-Service for the Cloud.*" CIDR 2011. (citations: 484)

[42] P. Cudre-Mauroux, E. Wu, S. Madden. "*TrajStore: An Adaptive Storage System for Very Large Trajectory Data Sets.*" ICDE 2010. (citations: 214)

[43] P. Cudre-Mauroux, E. Wu, S. Madden. "*The Case for RodentStore: An Adaptive, Declarative Storage System.*" CIDR 2009. (citations: 32)

[44] M. Cafarella, A. Halevy, D. Wang, E. Wu, Y. Zhang. "*WebTables: Exploring the Power of Tables on the Web.*" VLDB 2008. (acceptance rate: 16.8%, citations: 737)

[45] E. Wu, Y. Diao, S. Rizvi. "*High-performance complex event processing over streams.*" SIGMOD 2006. (acceptance rate: 13%, citations: 1138)

[46] M. J. Franklin, S. R. Jeffery, S. Krishnamurthy, F. Reiss, S. Rizvi, E. Wu, O. Cooper, A. Edakkunni, W. Hong. "*Design Considerations for High Fan-In Systems: The HiFi Approach.*" CIDR 2005. (citations: 268)

Short Papers, Technical Reports, and Demos

[47] L. Zhao, Q. Li, P. Wang, J. Wang, E. Wu. "*ActiveDeeper: A Model-based Active Data Enrichment system.*" VLDB 2020 demo.

[48] Y. Wu, R. Chang, J. Hellerstein, E. Wu. "*Facilitating Exploration with Interaction Snapshots under High Latency.*" InfoVIS (short paper) 2020.

[49] **L. Ramjit**, Z. Kong, R. Netravali, E. Wu. "*Physical Visualization Design.*" SIGMOD (demo) 2020.

[50] **L. Flokas**, Y. Wu, J. Wang, E. Wu. "*Towards Complaint-driven ML Workflow Debugging.*" MLOps 2020.

[51] S. Krishnan, E. Wu. "*AlphaClean: Automatic Generation of Data Cleaning Pipelines.*" ArXiv 2019. (citations: 12)

[52] **L. Arnett**, **R. Netzorg**, A. Chaintreau, E. Wu. "*Cross-platform Interactions and Popularity in the Live-streaming Community.*" CHI Latebreaking 2019.

[53] E. Wu. "*CIDR2: Crazier Innovations in Databases JOIN Reinforcement-learning Research.*" CIDR 2019 Abstract.

[54] **Y. Chen**, **Y. Shi**, B. Chen, **T. Sellam**, C. Vondrick, E. Wu. "*Deep Neural Inspection Using DeepBase.*" LearnSys 2018 Workshop at NIPS.

[55] **F. Psallidas**, E. Wu. "*Provenance in Interactive Visualizations.*" HILDA 2018.

[56] **H. Zhang**, **V. Rai**, **T. Sellam**, E. Wu. "*Precision Interfaces for Different Modalities.*" SIGMOD (demo) 2018.

[57] **F. Psallidas**, E. Wu. "*Demonstration of Smoke: A Deep Breath of Data-Intensive Lineage Applications.*" SIGMOD (demo) 2018.

[58] P. Wang, Y. He, R. Shea, J. Wang, E. Wu. "*Deeper: A Data Enrichment System Powered by Deep Web.*" SIGMOD (demo) 2018.

[59] **H. Zhang**, **T. Sellam**, E. Wu. "*Mining Precision Interfaces From Query Logs.*" Tech Report 2017.

[60] S. Krishnan, M. J. Franklin, K. Goldberg, E. Wu. "*BoostClean: Automated Error Detection and Repair for Machine Learning.*" Tech Report 2017. (citations: 36)

[61] M. Procopio, C. Scheidegger, E. Wu, R. Chang. "*Load-n-Go: Fast Approximate Join Visualizations That Improve Over Time.*" DSIA 2017.

[62] **G. Ryan**, A. Mosca, E. Wu, R. Chang. "*Approximate Entropy as a Measure of Line Chart Complexity.*" InfoVIS Poster 2017.

[63] Y. Wu, L. Xu, R. Chang, E. Wu. "*Towards a Bayesian Model of Data Visualization Cognition.*" DECISIVE 2017.

[64] **H. Nilforoshan**, J. Wang, E. Wu. "*PreCog: Improving Crowdsourced Data Quality Before Acquisition.*" Arxiv 2017.

[65] **H. Zhang**, **T. Sellam**, E. Wu. "*Precision Interfaces.*" HILDA 2017.

[66] S. Krishnan, E. Wu. "*PALM: Machine Learning Explanations For Iterative Debugging.*" HILDA 2017. (citations: 42)

[67] **H. Nilforoshan**, **J. Sands**, **K. Lin**, **R. Khanna**, E. Wu. "*Segment-Predict-Explain for Automatic Writing Feedback.*" Collective Intelligence 2017.

[68] **H. Nilforoshan**, **J. Sands**, **K. Lin**, **R. Khanna**, E. Wu. "*Dialectic: Enhancing Text Input Fields with Automatic Feedback to Improve Social Content Writing Quality.*" ArXiv 2017.

[69] E. Wu. "*CIDR: Chat-oriented Innovations in Database Research.*" CIDR 2017 Abstract.

[70] Y. Wu, J. Hellerstein, E. Wu. "*A DeVIL-ish Approach to Inconsistency in Interactive Visualizations.*" HILDA 2016.

[71] **D. Alabi**, E. Wu. "*PFunk-H: Approximate Query Processing using Perceptual Models.*" HILDA 2016. (citations: 21)

[72] S. Krishnan, D. Haas, M. J. Franklin, E. Wu. "*Towards Reliable Interactive Data Cleaning: A User Survey and Recommendations.*" HILDA 2016. (citations: 47)

[73] N. Kamat, E. Wu, A. Nandi. "*TrendQuery: A System for Interactive Exploration of Trends.*" HILDA 2016.

[74] S. Krishnan, M. Franklin, K. Goldberg, J. Wang, E. Wu. "*ActiveClean: An Interactive Data Cleaning Framework For Modern Machine Learning.*" SIGMOD 2016 Demo. (citations: 119)

[75] E. Wu, L. Jiang, L. Xu, A. Nandi. "*Graphical Perception in Animated Bar Charts.*" Arxiv 2016. (citations: 12)

[76] X. Wang, A. Meliou, E. Wu. "*QFix: Demonstrating error diagnosis in query histories.*" SIGMOD 2016 Demo.

[77] X. Wang, A. Meliou, E. Wu. "*QFix: Diagnosing errors through query histories.*" Arxiv 2016.

[78] S. Krishnan, J. Wang, E. Wu, M. J. Franklin, K. Goldberg. "*ActiveClean: Interactive Data Cleaning While Learning Convex Loss Models.*" Arxiv 2016.

[79] L. Battle, E. Benson, A. Parameswaran, E. Wu. "*Indexing Cost Sensitive Prediction.*" Technical Report 2016.

[80] E. Wu, A. Nandi. "*Towards Perception-aware Interactive Data Visualization Systems.*" DSIA 2015. (citations: 14)

[81] S. Krishnan, J. Wang, M. J. Franklin, K. Goldberg, T. Kraska, T. Milo, E. Wu. "*SampleClean: Fast and Reliable Analytics on Dirty Data (overview paper).*" IEEE Data Eng. Bulletin 2015. (citations: 34)

[82] D. Haas, S. Krishnan, J. Wang, M. J. Franklin, E. Wu. "*Wisteria: Nurturing Scalable Data Cleaning Infrastructure.*" VLDB 2015 demo. (citations: 39)

[83] A. Bhardwaj, A. Deshpande, A. Elmore, D. Karger, S. Madden, A. Parameswaran, H. Subramanyam, E. Wu, R. Zhang. "*Collaborative Data Analytics with Datahub.*" VLDB 2015 demo. (citations: 54)

[84] E. Wu, A. Marcus, S. Madden. "*Data In Context: Aiding News Consumers while Taming Dataspaces.*" DBCrowd 2013.

[85] E. Wu, S. Madden, M. Stonebraker. "*A Demonstration of DBWipes: Clean as You Query.*" VLDB 2012.

[86] C. Curino, E. Jones, Y. Zhang, E. Wu, S. Madden. "*Relational Cloud: The Case for a Database Service.*" MIT Tech Report 2010.

[87] E. Wu, P. Cudre-Mauroux, S. Madden. "*Demonstration of the TrajStore System.*" VLDB 2009 demo.

[88] M. Cafarella, N. Khoussainova, D. Wang, E. Wu, Y. Zhang, A. Halevy. "*Uncovering the Relational Web.*" WebDB 2008. (citations: 178)

[89] D. Gyllstrom, E. Wu, H. Chae, Y. Diao, P. Stahlberg, G. Anderson. "*SASE: Complex Event Processing over Streams (Demo).*" CIDR 2007. (citations: 276)

[90] D. Gyllstrom, E. Wu, H. Chae, Y. Diao, P. Stahlberg, G. Anderson. "*SASE: Complex Event Processing over Streams.*" CoRR 2006.

[91] M. N. Garofalakis, K. P. Brown, M. J. Franklin, J. M. Hellerstein, D. Z. Wang, E. Michelakis, L. Tancau, E. Wu, S. R. Jeffery, R. Aipperspach. "*Probabilistic Data Management for Pervasive Computing: The Data Furnace Project.*" IEEE Data Eng. Bulletin 2006. (citations: 41)

[92] O. Cooper, A. Edakkunni, M. J. Franklin, W. Hong, S. R. Jeffery, S. Krishnamurthy, F. Reiss, S. Rizvi, E. Wu. "*HiFi: A Unified Architecture for High Fan-in Systems.*" VLDB 2004 Demo. (citations: 39)

Primary Conference Names and 5 Year Impact Factor

VLDB	Very Large Databases. 3.56.
SIGMOD	SIG Management of Data. 3.41.
ICDE	IEEE Interational Conference on Data Engineering. 2.63.
CIDR	Conference on Innovative Data system Research. 3.3.
SOCC	ACM Symposium on Cloud Computing. 4.79.
TVCG	Transactions on Visualization and Computer Graphics. 4.558.
InfoVIS	Information Visualization. 4.558.
EuroVIS	Eurographics/IEEE Symposium on Visualization. 2.15
HILDA	Human in the Loop Data Analysis Workshop. 2.0
BuildSys	Conference on Systems for Energy-Efficient Built Environments. 2.02
DSIA	Data Systems for Interactive Analysis Workshop.
SysML	Conference on Machine Learning and Systems.