

El Kindi Rezig

CONTACT INFORMATION

50 S. Central Campus Dr.
Salt Lake City, UT 84112, USA

Phone: (801) 587-3206
E-mail: elkindi.rezig@utah.edu
Web: <https://elkindi.github.io>

RESEARCH INTERESTS

Data scientists spend most of their time doing “grunt” work, i.e., discovery, preparing and cleaning the data. The goal of my research is to build systems that target the main pain points in data science development: data discovery; data preparation and data debugging.

EDUCATION

- **Massachusetts Institute of Technology**, Cambridge, MA, USA
Postdoc, Computer Science and Artificial Intelligence Laboratory (CSAIL), August 2022
 - Advisor: Prof. Michael Stonebraker
- **Purdue University**, West Lafayette, IN, USA
Ph.D., Computer Science, August 2018
 - Advisors: Prof. Walid Aref and Prof. Mourad Ouzzani
- **University of Science and Technology Houari Boumediene (USTHB)**, Algiers, Algeria.
M.S., Computer Science, June 2010
 - Summa cum Laude
 - Advisor: Prof. Youcef Aklouf
- **University of Science and Technology Houari Boumediene (USTHB)**, Algiers, Algeria.
B.S., Computer Science, June 2008
 - Summa cum Laude

PUBLICATIONS

Double underline names are Ph.D. students as well as MS students whom I supervised at the University of Utah. Ph.D. students: Mahathir Mohammad and Arman Ashkari. MS students: Pavitra Mehra, Akash Khatri.

Single underline names are undergraduate students I supervised at the University of Utah.

Asterisk (*) denotes a visiting Ph.D. student whom I hosted in my lab for a semester.

- SIGMOD 2026 Yarden Gabbay, Haoquan Guan, Shaull Almagor, **El Kindi Rezig**, Brit Youngmann, Babak Salimi: Stress-Testing Causal Claims via Cardinality Repairs. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2026.
- SIGMOD 2026 Mahathir Mohammad, **El Kindi Rezig**: Qualitative Join Discovery in Data Lakes using Examples. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2026.
- SIGMOD 2026 Eugenie Lai, Inbal Croitoru, Brit Youngmann, Sainyam Galhotra, **El Kindi Rezig**, Michael Cafarella: Interpretable Attribute Discretization. *ACM SIGMOD International Conference on Management of Data*, 2026.
- CIDR 2026 **El Kindi Rezig**, Mahathir Mohammad, Nicolas Baret, Ricardo Mayerhofer, Andrew McNutt, Paul Rosen: Towards Scalable Visual Data Wrangling via Direct Manipulation. *Conference on Innovative Data Systems Research (CIDR)*, 2026.
- SIGMOD 2026 Mahathir Mohammad, **El Kindi Rezig**: Intent-aware Join Discovery using Natural Language [Demo]. *ACM SIGMOD International Conference on Management of Data*, 2026.
- SIGMOD 2026 Mahathir Mohammad, **El Kindi Rezig**: SemDisc: An End-to-End Query-by-Example Semantic Join Discovery System [Demo]. *ACM SIGMOD International Conference on Management of Data*, 2026.
- ICDE 2026 **El Kindi Rezig**: On breaking the scalability barrier in data cleaning [Abstract in the lightning talk track]. *IEEE International Conference on Data Engineering (ICDE 2026)*, 2026.

- SIGSPATIAL'25 **El Kindi Rezig**, Walid G. Aref: LORY: Location-aware Data Discovery from Data Lakes [Vision]. *ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL)*, 2025. [\[Best Vision Paper Award\]](#)
- VLDB'25 Akash Khatri, Mahathir Mohammad, **El Kindi Rezig**: Sort it Like You Mean It: Discovering Semantically Interesting Attribute Augmentations to Sort Tables [Demo]. *International Conference on Very Large Data Bases (VLDB)*, 2025.
- VLDB'25 Annabelle Warner, Andrew McNutt, Paul Rosen, **El Kindi Rezig**: Buckaroo: A Direct Manipulation Visual Data Wrangler [Demo]. *International Conference on Very Large Data Bases (VLDB)*, 2025.
- SIGMOD'25 Arman Ashkari, **El Kindi Rezig**: CausalExplain: Causal Explanations of Black-box Models with Training Data Subsets [Demo]. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2025.
- SIGMOD'25 Eugenie Lai, Inbal Croitoru, Noam Bitton, Ariel Shalem, Brit Youngmann, Sainyam Galhotra, **El Kindi Rezig**, Michael Cafarella: SeerCuts: Explainable Attribute Discretization [Demo]. *Proceedings of the ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2025.
- IEEE BigData'24 Arman Ashkari, **El Kindi Rezig**: Can Causal DAGs Help with Data-based Explanations of Black-box Models? *Data-centric AI Workshop @ IEEE International Conference on Big Data*, 2024
- IEEE BigData'24 Pavitra Mehra, **El Kindi Rezig**: Leveraging Structured and Unstructured Data to Clean Tables *Data-centric AI Workshop @ IEEE International Conference on Big Data*, 2024
- SIGCSE '23 Rowan Hart, Brian Hays, Connor McMillin, **El Kindi Rezig**, Gustavo Rodriguez-Rivera, Jeffrey A. Turkstra: Eastwood-Tidy: C Linting for Automated Code Style Assessment in Programming Courses. *ACM Technical Symposium on Computer Science Education (SIGCSE)*, 2023.
- CIDR'22 **El Kindi Rezig**, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Andrew Bowne, Lindsay McEvoy, Vijay Gadepally: Examples are All You Need: Iterative Data Discovery by Example in Data Lakes [Extended Abstract].
- Springer LNCS **El Kindi Rezig**, Vijay Gadepally, Timothy G. Mattson, Michael Stonebraker, Tim Kraska, Jun Kong, Gang Luo, Dejun Teng, Fusheng Wang: Heterogeneous Data Management, Polystores, and Analytics for Healthcare - VLDB Workshops, Poly 2022 and DMAH 2022. *Lecture Notes in Computer Science 13814*, Springer 2022, ISBN 978-3-031-23904-5.
- VLDB'21 **El Kindi Rezig**, Mourad Ouzzani, Walid Aref, Ahmed Elmagarmid and Ahmed R. Mahmood, Michael Stonebraker: Horizon: Scalable Dependency-driven Data Cleaning. *International Conference on Very Large Databases (VLDB)*, 2021.
- VLDB'21 **El Kindi Rezig**, Anshul Bhandari, Anna Fariha, Benjamin Price, Allan Vanterpool, Vijay Gadepally, Michael Stonebraker: DICE: Data Discovery by Example [Demo]. *International Conference on Very Large Databases (VLDB)*, 2021. .
- CIDR'21 **El Kindi Rezig**: Data Cleaning in the Era of Data Science: Challenges and Opportunities [Extended Abstract]. *Conference on Innovative Data Systems Research (CIDR)*, 2021
- arXiv **El Kindi Rezig**, Michael J. Cafarella, Vijay Gadepally: Technical Report on Data Integration and Preparation. CoRR abs/2103.01986 (2021)
- Poly@VLDB'20 **El Kindi Rezig**, Allan Vanterpool, Vijay Gadepally, Benjamin Price, Michael J. Cafarella, Michael Stonebraker: Towards Data Discovery by Example. *Workshop on Polystore systems for heterogeneous data in multiple databases (co-located with VLDB)*, 2020
- VLDB'20 **El Kindi Rezig**, Ashrita Brahmaroutu, Nesime Tatbul, Mourad Ouzzani, Nan Tang, Timothy G. Mattson, Samuel Madden, Michael Stonebraker: Debugging Large-Scale Data Science Pipelines using Dagger [Demo]. *International Conference on Very Large Databases (VLDB)*, 2020.
- CIDR'20 **El Kindi Rezig**, Lei Cao, Giovanni Simonini, Maxime Schoemans, Samuel Madden, Nan Tang, Mourad Ouzzani, Michael Stonebraker: Dagger: A Data (not code) Debugger. *Conference on Innovative Data Systems Research (CIDR)*, 2020.
- IEEE Data Eng. Bull.'19 Michael Stonebraker, **El Kindi Rezig**: Machine Learning and Big Data: What is Important? *IEEE Data Engineering Bulletin*, 2019

- VLDB'19 **El Kindi Rezig**, Lei Cao, Michael Stonebraker, Giovanni Simonini, Wenbo Tao, Samuel Madden, Mourad Ouzzani, Nan Tang and Ahmed Elmagarmid: Data Civilizer 2.0: A Holistic Framework for Data Preparation and Analytics. *International Conference on Very Large Databases (VLDB)*, 2019.
- HILDA@SIGMOD'19 **El Kindi Rezig**, Mourad Ouzzani, Ahmed Elmagarmid, Walid Aref, Michael Stonebraker: Towards an End-to-End Human-Centric Data Cleaning Framework. *Workshop on Human-In-the-Loop Data Analytics (co-located with SIGMOD)*, 2019
- arXiv **El Kindi Rezig**, Mourad Ouzzani, Walid Aref, Ahmed Elmagarmid and Ahmed R. Mahmood: Pattern-Driven Data Cleaning. CoRR abs/1712.09437 (2017)
- ICDE'16 **El Kindi Rezig**, Eduard Dragut, Mourad Ouzzani, Ahmed Elmagarmid and Walid Aref: ORLF: A Flexible Framework for Online Record Linkage and Fusion [Demo]. *IEEE International Conference on Data Engineering (ICDE)*, 2016.
- VLDB'15 Ahmed R. Mahmood, Ahmed M. Aly, Thamir Qadah, **El Kindi Rezig**, Anas Daghistani, Amgad Madkour, Ahmed S. Abdelhamid, Mohamed S. Hassan, Walid G. Aref, Saleh Basalamah: Tornado: A Distributed Spatio-Textual Stream Processing [Demo]. *International Conference on Very Large Databases (VLDB)*, 2015.
- ICDE'15 **El Kindi Rezig**, Eduard C. Dragut, Mourad Ouzzani, Ahmed K. Elmagarmid: Query-Time Record Linkage and Fusion over Web Databases. *IEEE International Conference on Data Engineering (ICDE)*, 2015.
- JBIS'13 **El Kindi Rezig**, Youcef Aklouf, Hadj Madani Meghazi: Leveraging human experts' knowledge to detect and publish compositions of Semantic Web services in a repository. *International Journal of Business Information Systems*, 2013
- SIGMOD'11 Hazem Elmeleegy, Jaewoo Lee, **El Kindi Rezig**, Mourad Ouzzani, Ahmed Elmagarmid: UMAP: A System for Usage-Based Schema Matching and Mapping [Demo]. *ACM Special Interest Group on Management of Data Conference (SIGMOD)*, 2011.
- ICEIS'10 **El Kindi Rezig**, Valerie Monfort: Towards Location-Based Services standardization: An Application based on Mobility and Geo-Location. *International Conference on Enterprise Information Systems*, 2010
- IGI Global (book chapter) Youcef AKLOUF , **El Kindi Rezig**: Rule-based approach for a better B2B discovery. *Organizational Advancements through Enterprise Information Systems: Emerging Applications, Chapter 19, IGI Global, 2009, USA*
- PATENTS Paolo Papotti, Felix Naumann, Sebastian Kruse, **El Kindi Rezig**: Systems and Methods for Data Integration. Patent No 20160154830. June 2016.
- EDITED BOOKS **El Kindi Rezig** , Vijay Gadepally, Timothy G. Mattson, Michael Stonebraker, Tim Kraska, Fusheng Wang, Gang Luo, Jun Kong, Alevtina Dubovitskaya: Heterogeneous Data Management, Polystores, and Analytics for Healthcare - VLDB Workshops, Poly 2021 and DMAH 2021, Revised Selected Papers. ISSN 0302-9743. Springer Lecture Notes in Computer Science.
- Joint Proceedings of Workshops at the 49th International Conference on Very Large Data Bases (VLDBW 2023), Vancouver, Canada, August 28 - September 1, 2023. Edited by: Rajesh Bordawekar, Cinzia Cappiello, Vasilis Efthymiou, Lisa Ehrlinger, Vijay Gadepally, Sainyam Galhotra, Sandra Geisler, Sven Groppe, Le Gruenwald, Alon Halevy, Hazar Harmouch, Oktie Hassanzadeh, Ihab Ilyas, Ernesto Jiménez-Ruiz, Sanjay Krishnan, Tirthankar Lahiri, Guoliang Li, Jiaheng Lu, Wolfgang Mauerer, Umar Farooq Minhas, Felix Naumann, M. Tamer Özsu, **El Kindi Rezig**, Kavitha Srinivas, Michael Stonebraker, Satyanarayana R Valluri, Maria-Esther Vidal, Haixun Wang, Jiannan Wang, Yingjun Wu, Xun Xue, Mohamed Zait, Kai Zeng Submitted by: Jiannan Wang Published on CEUR-WS: 22-Aug-2023.
- GRANTS
- **Training the Next Generation of Database Engineers: Responsible, User-Centric, and Efficient Data Management.** USHE: Computer Science Targeted Workforce Application. Academic year 2023-2024. Grant for developing a new course: Manage Data for & with Machine Learning.

- **NSF: National Artificial Intelligence (AI) Research Institutes: NSF & Simons Institute for Artificial Intelligence and Cosmic Origins (NAICO): A Nexus for Accessible Astronomy** [Funded Personnel, PI: Stella Offner, University of Texas at Austin]. Amount: \$20,000,000. Share: \$400,000.
- The University of Utah Faculty Small Grants Program (VPR Office): **Data-based explanations of black box algorithms**. Role: Sole PI. Amount: \$10,000. Spring 2024.

AWARDS AND HONORS

- Best Vision Paper Award, ACM SIGSPATIAL 2025.
- Raymond Boyce teaching award (highest teaching award in the Purdue CS dept.), 2017.
- Purdue Computer Science Teaching Fellowship, 2015 - 2016.
- Purdue Graduate School Summer Research Grant, 2015.
- Purdue CS travel award, 2016.
- ICDE Student Travel Award, 2015.
- Purdue College of Science Graduate Student International Travel Award, 2015.
- Ranked 1st at the Microsoft Imagine Cup competition, Software Design category, Algeria, 2009.
- Microsoft Student Partner, USTHB, 2009 - 2010.
- Top computer science student award (most prestigious university-wide student award given to the first-ranking student in the M.S. program), USTHB, Algeria, 2010.

TEACHING
EXPERIENCE

University of Utah

Term	Course name
Fall 2023	CS3190: Foundations of Data Analysis
Fall 2023	DS3190: Foundations of Data Analysis
Fall 2023	COMP5190: Foundations of Data Analysis
Spring 2024	CS4964: Manage Data for & with ML (new course)
Spring 2024	CS6964: Manage Data for & with ML (new course)
Fall 2024	CS3190: Foundations of Data Analysis
Fall 2024	DS3190: Foundations of Data Analysis
Fall 2024	COMP5190: Foundations of Data Analysis
Spring 2025	CS4964: Manage Data for & with ML (new course)
Spring 2025	CS6964: Manage Data for & with ML (new course)
Fall 2025	CS3190: Foundations of Data Analysis
Fall 2025	DS3190: Foundations of Data Analysis
Fall 2025	COMP5190: Foundations of Data Analysis
Spring 2026	CS4964: Manage Data for & with ML (new course)
Spring 2026	CS6964: Manage Data for & with ML (new course)

CS4964/CS6964 is a special topics course that I developed.

Purdue University

- (Summer 2017 and 2018) CS50011: Introduction To Systems For Information Security. Role: **Co-developer and instructor (2 terms)**.
- (8 terms) CS252: Systems Programming. Role: TA (6 terms) and **instructor (2 terms)**.
- (Spring 2018) CS251: Algorithms and Data Structures. Role: TA (1 term).
- (Spring 2017) CS240: Programming in C. Role: TA (1 term).

MENTORSHIP

- I am currently advising three Ph.D. students and one undergraduate student. Previously, I have worked with Master's students and undergraduate students at the University of Utah.
 - Mir Mahathir Mohammad (Ph.D. student) (since Fall 2023)
 - Arman Ashkari (Ph.D. student) (since Fall 2023)
 - Rabeya Hossain (Ph.D. student) (since Fall 2025)
 - Annabelle Warner (undergraduate student co-advised with Paul Rosen and Andrew McNutt) (Spring 2025). **Received the KSoC Best Undergraduate Research award in Spring 2025 for our project.**
 - Nicolas Baret (undergraduate student, co-advised with Paul Rosen and Andrew McNutt), Summer 2025.
 - Jacob Ormsby (undergraduate student) (since Spring 2025)
 - Adeel Aslam (visiting Ph.D. student) (Fall 2024).
 - Pavitra Mehra (master's student) (Fall 2023- Spring 2025)
 - Akash Khatri (master's student) (Fall 2024)
 - Abhishek Rajgaria (master's student) (Summer 2024)
 - Zane Colter (undergraduate student) (Spring 2024)
 - William Erignac (undergraduate student) (Spring 2024, Summer 2024, Fall 2024)
- Mentored 6 students (4 MIT Master of Engineering/graduate, 1 undergraduate and 1 high school), MIT, 2018 - 2022
- Mentored Master Theses
 - Peter Griggs: Database updates using interactive Pan and Zoom visualizations. MIT EECS, 2020
 - Zhaoyuan Zhang: A New Authoring System for Diverse Data Visualization At Scale. MIT EECS, 2021
 - Jim Peraino: Architectural epidemiology: a computational framework. MIT EECS, 2020
 - Erica Zhou: Interactive visualization and discovery of possible transmission routes of Clostridioides difficile. MIT EECS, 2020

PROFESSIONAL EXPERIENCE

The University of Utah - School of Computing, Salt Lake City, UT, USA.

Assistant Professor

August 2023 to now

Massachusetts Institute of Technology - CSAIL, Cambridge, MA, USA.

Research Scientist

October 2022 to July 2023

Massachusetts Institute of Technology - CSAIL, Cambridge, MA, USA.

Postdoctoral Associate

November 2018 to October 2022

Purdue University, West Lafayette, IN, USA.

Graduate Teaching Assistant / Graduate Instructor

June 2014 to August 2018

Graduate Research Assistant

August 2010 to May 2014

Qatar Computing Research Institute, Doha, Qatar.

Research Associate Intern

Jan 2012 to Jul 2012 and May 2013 to Aug 2013

INEODEV Ltd, Algiers.

Software developer

July 2008 to July 2010

SERVICE

- Member of the Best Demo Committee at SIGMOD 2025.
- Session chair at SIGMOD 2025, Berlin.
- Session chair at VLDB 2025, London.
- Chair of a VLDB workshop (Poly@VLDB), 2023.
- PC chair of the North East Database Day conference (NEDB), 2023
- Reviewer of SIGMOD 2026, SIGMOD 2025, VLDB 2025, ICDE 2025, SIGMOD 2023, VLDB 2024, SIGKDD 2023, and TKDE 2023
- Reviewer of the SIGMOD Record 2021
- Session chair at SIGKDD 2021 (research track)
- PC member of SIGMOD 2022 (demo track)
- Chair of the Poly'21 and Poly'22 workshops (co-located with VLDB 2021 and VLDB 2022)
- Reviewer of: SIGMOD 2023, HILDA@SIGMOD 2022, SIGMOD 2022 (Demo), VLDB 2022 (Demo), HILDA@SIGMOD 2022, SIGKDD 2021, SIGKDD 2022, DASFAA 2021, the VLDB Journal 2021, ACM Journal of Data and Information Quality 2021, SIAM International Conference on Data Mining 2020, IEEE Transactions on Knowledge and Data Engineering. 2018 - 2020
- Demo session chair at VLDB 2020
- Volunteer student organizer at the ICDE conference 2014 (Chicago) and 2016 (Helsinki)
- External reviewer: CIKM (2015), SSTD (2015), TKDE (2014), EDBT (2013, 2014), IEEE Transactions on Services Computing (2011), WISE (2011)
- After-school Math and English tutoring for at-risk youth at the Lafayette Transitional Housing Center, Lafayette, IN, USA, 2013
- Vice president of the .NET Club at the USHTB computer science department, Algeria, 2010

INTERNAL SERVICE

- Instructor for the Faculty Upskilling in AI (UAI) Program, a program designed to equip University of Utah faculty across departments with AI concepts and tools they can use in their own research, Spring 2025.
- Participated in Phase 3 of the Faculty Upskilling in AI (UAI) program, where I co-led (with Vivek) a panel to discuss concrete problems from the AUI participants. July 2025.
- Co-instructor of the Data Science Seminar, Fall 2025.
- Director of the Graduate Data Science Certificate program, 2025.
- Member of the graduate admission committee, 2024/2025.
- Teaching Area Coordinator for the Data Management and Analysis area (since 12/23)
- Speaker at the Grad Visit Day to introduce Data Management research at UtahDB.
- Faculty Advisor of the ACM club (the club died down during COVID, so I resurrected it and had to attract students and form a new officers board). The club has been active since Spring 2024.
- Co-founder of the UtahDB research center, The University of Utah
- Capstone projects judge (Fall 2023)
- Mentor of a first-year Ph.D. student

INVITED TALKS

- Invited talk at the DaSH (Workshop on Data Science with Human in the Loop) VLDB workshop, 2025, London. Talk title: Unleashing Data Science: It's Time to Fix the Data Preparation Problem. September 2025.
- Invited panelist at the DaSH (Workshop on Data Science with Human in the Loop) VLDB workshop, 2025, London. September 2025.
- Invited talk at Hopara, Inc. (a company specializing in scalable data visualizations). Talk title: Buckaroo: A Direct Manipulation Data Wrangler. May 2025.
- Invited talk at Tamr, Inc. (a company specializing in data integration). Talk title: Buckaroo: A Direct Manipulation Data Wrangler. May 2025.
- Recent Trends in Scalable and Human-centric Data Systems. Research Challenges in Computing Seminar Series, University of Utah, October 2023.
- Data Preparation: The Biggest Roadblock in Data Science. Data Science Seminar series, University of Utah, October 2023.
- Dagger: a Provenance-Based Data Debugging System. Intel Labs. March 2021
- DICE: Data Discovery by Example. AI Accelerator annual meeting. April 2021
- Dagger: a Provenance-Based Data Debugging System. Data Systems and Artificial Intelligence (DSAIL) convention at MIT. April 2021
- Data Cleaning in the Era of Data Science: Challenges and Opportunities. CIDR. January 2020
- Debugging Large-Scale Data Science Pipelines using Dagger. VLDB. August 2020
- Dagger: a Data (not code) Debugger. CIDR. January 2020
- Data debugging for Data Science. Qatar Computing Research Institute. March 2019