

# Zhengjie Miao

*Ph.D. in Computer Science*

444 Castro Street Suite 720  
Mountain View, CA 94041

✉ [zhengjie@megagon.ai](mailto:zhengjie@megagon.ai)

🌐 [www.miaozhengjie.com](http://www.miaozhengjie.com)

🔍 Google Scholar

## Research Interests

I am interested in improving human-data interaction in data science pipelines, which broadly spans data management, machine learning, and visual analytics. Currently my research focuses on helping users:

- write and debug relational queries
- explain analytical query results
- integrate and prepare their data using machine learning techniques

## Education

9/2022 **Ph.D. in Computer Science**, *Duke University*, Durham NC, *GPA: 3.9/4.0*.

- Thesis: Simplifying Human-in-the-loop Data Science Pipeline: Explanation, Debugging, and Data Preparation
- Advisor: Sudeepa Roy

12/2016 **M.S. in Computer Science**, *Columbia University*, New York NY, *GPA: 4.0/4.0*.

- Advisor: Eugene Wu

7/2015 **B.S. in Computer Science and Technology**, *Peking University*, Beijing - China, *GPA: 3.5/4.0*.

- Advisor: Xiaoru Yuan

## Professional Experience

9/2022-  
Now **Research Scientist**, *Megagon Labs*.

5/2021-  
8/2021 **Research Intern**, *Microsoft Research (Data Systems Group)*, supervised by Dr. Yeye He.

5/2020-  
8/2020 **Research Intern**, *Megagon Labs*, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan.

5/2019-  
8/2019 **Research Intern**, *Megagon Labs*, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan.

## Awards

2022 Outstanding Ph.D. Dissertation, *Department of Computer Science, Duke University*

2020 Microsoft Research PhD Fellowship Finalist

2019 Outstanding Ph.D. Research Initiation Project Award, *Department of Computer Science, Duke University*

2019 VLDB Travel Grant

2018, 2019 ACM SIGMOD Travel Award

2015 7th Place in ACM/ICPC Greater New York Regional

2014 Award for Excellent Detailed Analysis, *IEEE Visual Analytics Science and Technology (VAST) Challenge Mini-Challenge 1*

2013 The May Fourth Scholarship, *Peking University*

2012 Silver medal in ACM/ICPC Asia Regional Contest in Tianjin

2009 First Prize in National Olympiad in Informatics in Hunan Province (NOIP)

---

## Research Full Papers

- SIGMOD 2022 **Understanding Queries by Conditional Instances**, Amir Gilad\*, Zhengjie Miao\*, Sudeepa Roy, and Jun Yang, \* denotes equal contribution, Proceedings of the 2022 ACM SIGMOD International Conference on Management of Data, June 2022.
- SIGMOD 2021 **Rotom: A Meta-Learned Data Augmentation Framework for Entity Matching, Data Cleaning, Text Classification, and Beyond**, Zhengjie Miao, Yuliang Li, and Xiaolan Wang, Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data, June 2021.
- SIGMOD 2021 **Putting Things into Context: Rich Explanations for Query Answers using Join Graphs**, Chenjie Li, Zhengjie Miao, Qitian Zeng, Boris Glavic, and Sudeepa Roy, Proceedings of the 2021 ACM SIGMOD International Conference on Management of Data, June 2021.
- WWW 2020 **Snippext: Semi-supervised Opinion Mining with Augmented Data**, Zhengjie Miao, Yuliang Li, Xiaolan Wang, and Wang-Chiew Tan, Proceedings of The Web Conference, April 2020.
- SIGMOD 2019 **Explaining Wrong Queries Using Small Examples**, Zhengjie Miao, Sudeepa Roy, and Jun Yang, Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data, June 2019.
- SIGMOD 2019 **Going Beyond Provenance: Explaining Query Answers with Pattern-based Counterbalances**, Zhengjie Miao\*, Qitian Zeng\*, Boris Glavic, and Sudeepa Roy, \* denotes equal contribution, Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data, June 2019.
- CIDR 2017 **Combining Design and Performance in a Data Visualization Management System**, Eugene Wu, Fotis Psallidas, Zhengjie Miao, Haoci Zhang, Laura Rettig, Yifan Wu, and Thibault Sellam, 8th Biennial Conference on Innovative Data Systems Research, Jan 2017.

---

## Short/Demonstration Papers

- SIGMOD 2023 **Characterizing and Verifying Queries Via CinsGen: Explaining Wrong Relational Queries Using Small Examples**, Hanze Meng, Zhengjie Miao, Amir Gilad, Sudeepa Roy, and Jun Yang, Proceedings of the 2023 ACM SIGMOD International Conference on Management of Data (to appear).
- VLDB 2022 **Cajade: Explaining query results by augmenting provenance with context**, Chenjie Li, Juseung Lee, Zhengjie Miao, Boris Glavic, and Sudeepa Roy, Proceedings of the VLDB Endowment, Vol. 15, No. 12.
- VLDB 2021 **Data Augmentation for ML-driven Data Preparation and Integration (Tutorial)**, Yuliang Li, Xiaolan Wang, Zhengjie Miao, and Wang-Chiew Tan, Proceedings of the VLDB Endowment, Vol. 14, No. 12.
- VLDB 2020 **I-Rex: An Interactive Relational Query Explainer for SQL**, Zhengjie Miao, Tiangang Chen, Alexander Bendeck, Kevin Day, Sudeepa Roy, and Jun Yang, Proceedings of the VLDB Endowment, Vol. 13, No. 12.
- VLDB 2019 **CAPE: Explaining Outliers by Counterbalancing**, Zhengjie Miao\*, Qitian Zeng\*, Chenjie Li, Boris Glavic, Oliver Kennedy, and Sudeepa Roy, \* denotes equal contribution, Proceedings of the VLDB Endowment, Vol. 12, No. 12.
- VLDB 2019 **LensXPlain: Visualizing and Explaining Contributing Subsets for Aggregate Query Answers**, Zhengjie Miao, Andrew Lee, and Sudeepa Roy, Proceedings of the VLDB Endowment, Vol. 12, No. 12.
- SIGMOD 2019 **RATest: Explaining Wrong Relational Queries Using Small Examples**, Zhengjie Miao, Sudeepa Roy, and Jun Yang, Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data, June 2019.

---

## Teaching and Professional Services

- TA CompSci 316: Introduction to Database Systems, Spring 2019
- TA CompSci 216: Everything Data, Spring 2018
- PC Member Proceedings of the VLDB Endowment, Vol. 17
- PC Member ACM SIGMOD International Conference on Management of Data, 2024

PC Member Workshop on Human-In-the-Loop Data Analytics (HILDA), 2023  
PC Member International Conference on Data Engineering, 2023  
Reviewer Journal of Data and Information Quality, 2023  
Reviewer International Conference on Data Engineering, 2022  
Reviewer International Conference on Database Theory, 2021  
Reviewer ACM Transactions on Database Systems, 2020  
PC Member Proceedings of the VLDB Endowment Reproducibility, 2019