

Zhengjie Miao

Ph.D. in Computer Science

8888 University Drive, TASC 1 9407
Burnaby, BC, Canada V5A 1S6

✉ zhengjie@sfu.ca

🌐 www.miaozhengjie.com

🔍 [Google Scholar](#)

Education

- 2017-2022 **Ph.D. in Computer Science**, *Duke University*, Durham NC, *GPA: 3.9/4.0*.
 - Dissertation: Explanations in the Data Science Pipeline
 - Advisor: Sudeepa Roy
- 2015-2016 **M.S. in Computer Science**, *Columbia University*, New York NY, *GPA: 4.0/4.0*.
 - Advisor: Eugene Wu
- 2011-2015 **B.S. in Computer Science and Technology**, *Peking University*, Beijing - China, *GPA: 3.5/4.0*.
 - Advisor: Xiaoru Yuan

Professional Experience

- 10/2023-
Now **Assistant Professor**, *Simon Fraser University, School of Computing Science*, Burnaby, BC.
- 9/2022-
10/2023 **Research Scientist**, *Megagon Labs*, Mountain View, CA.
- 5/2021-
8/2021 **Research Intern**, *Microsoft Research (Data Systems Group)*, supervised by Dr. Yeye He, Durham, NC (remote).
- 5/2020-
8/2020 **Research Intern**, *Megagon Labs*, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan, Durham, NC (remote).
- 5/2019-
8/2019 **Research Intern**, *Megagon Labs*, supervised by Dr. Yuliang Li and Dr. Wang-Chiew Tan, Mountain View, CA.

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, *Duke University*
- 2019 VLDB Travel Grant
- 2018, 2019 ACM SIGMOD Travel Award
- 2015 7th Place in ACM/ICPC Greater New York Regional
- 2013 The May Fourth Scholarship, *Peking University*

Research Full Papers

- SIGMOD 2024 **Watchog: A Light-weight Contrastive Learning based Framework for Table Understanding**, *Zhengjie Miao, Jin Wang*.
ACM SIGMOD International Conference on Management of Data, to appear
- 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*, Link to [📄](#).
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*, Link to [📄](#).
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*, Link to [📄](#).
The Web Conference 2020, April 2020
- SIGMOD 2019 **Explaining Wrong Queries Using Small Examples**, *Zhengjie Miao, Sudeepa Roy, and Jun Yang*, Link to [📄](#).
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, Link to [📄](#).
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, Thibault Sellam*, Link to [📄](#).
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, June 2023.
- 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*, Link to [📄](#).
Proceedings of the VLDB Endowment (PVLDB), 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*, Link to [📄](#).
Proceedings of the VLDB Endowment (PVLDB), 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, Link to [📄](#).
Proceedings of the VLDB Endowment (PVLDB), Vol. 12, No. 12
- VLDB 2019 **LensXPlain: Visualizing and Explaining Contributing Subsets for Aggregate Query Answers**, *Zhengjie Miao, Andrew Lee, and Sudeepa Roy*, Link to [📄](#).
Proceedings of the VLDB Endowment (PVLDB), Vol. 12, No. 12
- SIGMOD 2019 **RATest: Explaining Wrong Relational Queries Using Small Examples**, *Zhengjie Miao, Sudeepa Roy, and Jun Yang*, Link to [📄](#).
ACM SIGMOD International Conference on Management of Data, June 2019

Teaching, Mentoring and Other Services

- Mentor M.S. project, Qiulin Li, 2021
- Mentor M.S. project, Tiangang Chen, Kehan Lyu, 2020
- Mentor Duke undergraduate CS+: CompSci Projects Beyond the Classroom, Alexander Bendeck (now PhD student at Georgia Tech), Kevin Day, Jeffrey Luo, 2020
- Mentor Undergraduate independent study, Jeremy Cohen, 2020

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 International Conference on Data Engineering, 2022
Reviewer International Conference on Database Theory, 2021
Reviewer ACM Transactions on Database Systems, 2020
Volunteer ACM SIGMOD International Conference on Management of Data, 2020
PC Member Proceedings of the VLDB Endowment Reproducibility, 2019