

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.
◦ Dissertation: Simplifying Human-in-the-loop Data Science Pipeline: Explanations, Debugging, and Data Preparation
◦ Advisor: Sudeepa Roy
- 2015-2016 **M.S. in Computer Science**, *Columbia University*, New York, NY.
◦ Advisor: Eugene Wu
- 2011-2015 **B.S. in Computer Science and Technology**, *Peking University*, Beijing, China.
◦ Advisor: Xiaoru Yuan

Professional Experience

- 10/2023- Present **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, *Department of Computer Science, Duke University*
- 2019 VLDB Travel Grant
- 2018, 2019 ACM SIGMOD Travel Award
- 2013 The May Fourth Scholarship, *Peking University*

Research Full Papers

- CHI 2024 **Human-LLM Collaborative Annotation Through Effective Verification of LLM Labels**, *Xinru Wang, Hannah Kim, Sajjadur Rahman, Kushan Mitra, Zhengjie Miao*.
Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (to appear)
- SIGMOD 2024 **Watchdog: A Light-weight Contrastive Learning based Framework for Table Understanding**, *Zhengjie Miao, Jin Wang*, Link to [🔗](#).
Proc. ACM Manag. Data 1, 4, Article 272 (December 2023)

- SIGMOD 2022 **Understanding Queries by Conditional Instances**, Amir Gilad*, Zhengjie Miao*, Sudeepa Roy, and Jun Yang, * denotes equal contribution, Link to [📄](#).
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, Link to [📄](#).
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

Mentoring

- 2024– M.S. Thesis Student@SFU: Xinyi Huang
- Summer 2023 Research Intern@Megagon Labs: Xinru Wang (Purdue University), co-mentored with Hannah Kim

- Summer 2023 Research Intern@Megagon Labs: Nima Shahbazi (University of Illinois Chicago), co-mentored with Jin Wang, Nikita Bhutani
- 2020-2021 M.S. Project@Duke: Qiulin Li
- 2019-2020 M.S. Project@Duke: Tiangang Chen
- 2019-2020 M.S. Project@Duke: Kehan Lyu
- 2020 Duke Undergraduate CS Plus: Alexander Bendeck (now PhD student at Georgia Tech), Kevin Day, Jeffrey Luo
- 2020 Undergraduate Independent Study@Duke: Jeremy Cohen

Teaching

- Spring 2024 **Instructor**, *CMPT 459/984: Special Topics in Database Systems*, SFU.
- Spring 2019 **TA**, *CompSci 316: Introduction to Database Systems*, Duke University.
- Spring 2018 **TA**, *CompSci 216: Everything Data*, Duke University.

Professional Services

Program Committee Member

- 2025 SIGMOD
- 2024 Workshop on Human-In-the-Loop Data Analytics (HILDA)
- 2024 VLDB
- 2024 SIGMOD
- 2023 Workshop on Human-In-the-Loop Data Analytics (HILDA)
- 2023 ICDE
- 2023 ICDE (demonstration track)
- 2019 VLDB Reproducibility

Reviewer

- 2023 ACM Journal of Data and Information Quality
- 2020 ACM Transactions on Database Systems

External Reviewer

- 2022 ICDE
- 2021 ICDT