

# YUE GONG

✉ yuegong@uchicago.edu · ☎ (+1) 765-337-4319 · 🌐 snowgy · 🏠 homepage

## EDUCATION

---

**The University of Chicago** 2020 - 2025(expected)

*Ph.D.* in Computer Science, *Advisor:* Raul Castro Fernandez

*Research Area:* Data discovery/exploration

**Southern University of Science and Technology** Sep. 2016 - July 2020

*Bachelor* in Computer Science and Engineering

GPA: 3.88/4.00, *Awarded as Summa Cum Laude*

## PUBLICATION

---

**Ver: View Discovery in the Wild.** *Yue Gong\**, *Zhiru Zhu\**, *Sainyam Galhotra*, *Raul Castro Fernandez*. *IEEE International Conference on Data Engineering (ICDE) 2023*. [pdf] [code] \* equal contribution

**Metam: Goal-Oriented Data Discovery.** *Sainyam Galhotra*, *Yue Gong*, *Raul Castro Fernandez*. *IEEE International Conference on Data Engineering (ICDE) 2023*. [code]

## WORK EXPERIENCE

---

**Research Intern, TikTok Infrastructure Lab** June. 2022 - Sep. 2022

- Implemented Analyze Table command in an HSAP(Hybrid serving/analytics processing) system
- Investigated STOA distinct count algorithms and their implementations in Clickhouse and Apache Doris
- Proposed a solution to collect stats incrementally in a query optimizer.

**Software Engineering Intern, Tencent Ads** April. 2020 - July. 2020

- Participate in building an internal tool for managing and monitoring Ads data with Tencent Ads team.
- Developed RESTful APIs for manipulating Ads data stored in MySQL using gRPC and Gorm framework.

**Research Intern, University of California, Irvine,**[poster] [report] July. 2019 - Sep. 2019

- Optimized query execution by pushing down operators in query plans on an open-source database AsterixDB.
- Explored three query patterns where the optimization using pushdown can have an up to 34.6% speed-up.

## PROJECT

---

**Ver: View Discovery in the Wild** Oct. 2020 - Present

*The University of Chicago* [paper] [demo]

- Built **Ver**, a data discovery system that identifies project-join views over large repositories of tables that do not contain join path information, and even when input queries are inaccurate.
- Implements novel column selection algorithm robust to noisy input and efficient join-graph-search module
- Demonstrate its performance with large-scale end-to-end experiments on real-world datasets containing tens of millions of join paths.

**METAM: Goal-Oriented Data Discovery, The University of Chicago** Sep. 2021 - Present

- METAM is a novel goal-oriented framework that queries the downstream task with a dataset candidate, forming a feedback loop that automatically steers the discovery and augmentation process.
- Implemented the join path search module and extracted signals from join paths to help profile datasets.