

# MINGHAO YAN

Email: [Minghao.Yan@rice.edu](mailto:Minghao.Yan@rice.edu)

Personal Website: <https://minghaoyan.netlify.app/>

---

## EDUCATION

**Master of Science in Computer Science,**

**Expected: May 2022**

**Bachelor of Science in Computer Science, Summa Cum Laude**

**May 2021**

**Rice University, Houston, Texas**

GPA: 4.03/4.09 (Only the first year was weighted due to school policy. Top 5% in the School of Engineering)

**Relevant Courses:**

**Math/Stats:** Honors Linear Algebra; Analysis I; Linear Regression; Operation Research; Information Theory.

**System/Arch/PL:** Database Systems; Principles of Parallel Programming; Advanced Object-Oriented Design; Principles of Programming Languages; Operating Systems and Concurrent Programming; Computer Architecture; Software Engineering Methodology.

**AI/ML/Algorithms:** Artificial Intelligence; Machine Learning; Deep Learning; Reasoning about Algorithms; Probabilistic Algorithms and Data Structures.

## EXPERIENCE

**Research Assistant at Rice University Hashing and Sketching Lab, May 2019 – Now**

**Advised by Prof. Anshumali Shrivastava**

- Current Project: Design and implementation of a MPI based distributed framework for fast Neural Network training on CPUs by leveraging hashing techniques (**Manuscript in preparation**).
- Previous Project: Designed and implemented a fast Bloom Filter based algorithm for Multiple Set Membership Testing (MSMT) problem that can index 170TB of genomic data in less than 9 hours and query in sublinear time (**Published at SIGMOD 2021**).

**Research Assistant in The Rice Undergraduate Data Science Summer Program**

**(RUDSSP), May 2018 – July 2018**

**Advised by Prof. Luay Nakhleh, Prof. Marina Vanucci, and Prof. Hamim Zafar**

**Project: Simulating, inferring, and scaling up single nucleotide variants in single-cell genomic data**

- Built a genome data simulator for ground truth generation
- Devised and implemented a genome amplifier
- Optimized a genome reads generator

## PUBLICATION

- Fast Processing and Querying of 170TB of Genomics Data via a Repeated And Merged BloOm Filter (RAMBO)

**Gaurav Gupta\*, Minghao Yan\***, Benjamin Coleman, Bryce Kille, R. A. Leo Elworth, Tharun Medini, Todd Treangen, Anshumali Shrivastava

**SIGMOD 2021.**

(\* denotes equal contribution)

## AWARDS

- Summa Cum Laude (Top 5% in the School of Engineering), 2021
- President's Honor Roll, 6/8 Semesters
- Research Fellowship for Master of Science in Computer Science, 2020
- Louis J. Walsh Scholarship, 2019-2020, 2020-2021
- RUDSSP Summer Fellowship (Top 11 among 60+ candidates), May - Jul 2018

## **TEACHING EXPERIENCE (TA)**

- Fall 2018: Honors Calculus III (MATH221)
- Spring 2019: Algorithmic Thinking (COMP182)
- Fall 2019, 2020: Reasoning About Algorithms (COMP382)
- Spring 2020: Honors Calculus IV (MATH222)

## **SKILLS**

- Languages: English (Bilingual); Mandarin (Native)
- Proficient with Python, Java, and C++; Basics of MATLAB, R, C#, and JavaScript.