Hao Jiang, Ph.D.

Postdoctoral Fellow
Harvard John A. Paulson
School Of Engineering And Applied Sciences
150 Western Ave, Boston, MA 02134

(315)267-6271
hajiang@seas.harvard.edu
https://harperjiang.github.io

I am an accomplished Postdoctoral Researcher at Harvard University. My research involves self-designing blockchain and large language models. They have many potential applications in a variety of industries. Previously, I earned my Ph.D. in Computer Science from the University of Chicago, where I specialized in Databases, Distributed Systems, and Machine Learning. I have authored several publications in leading academic conferences. With six years of experience as a Software Engineer, I am well-versed in Java, C++, and Python. I also have experience leading small teams.

Research and Work Experience

- **Harvard University** — Boston, MA
  Postdoctoral Researcher with Prof. Stratos Idreos — *Oct 2021 – Now*
  - Build a self-designing blockchain system that adapts to workloads to provide high-throughput and low-latency
  - Develop a small and detachable memory component for large language models. These components maintain domain knowledge and long-term memory and can be trained independently from the main LLM. They grant the LLM different abilities with low cost.
- **Facebook Inc.** — Menlo Park, CA
  Software Engineer Intern — *July. 2018 – Sept. 2018*
  - Design, implement and evaluate a load balance algorithm for PHP requests
- **The University of Chicago** — Chicago, IL
  Research Assistant under Prof. Aaron J. Elmore — *Sept. 2015 – Aug. 2021*
  - Develop a Compression-Aware Query Engine for Columnar Database that has 5x performance compared to the state-of-the-art. Result published on SIGMOD.
  - Design a neural network based classifier to choose the best compression algorithm for a given dataset.
  - Develop an innovative, fast, and efficient compression algorithm for String data. The algorithm compresses better and faster than the previous state-of-the-art. Result published on VLDB.
  - Develop a fast query algorithm on compressed data without decompression. It queries compressed data 20-100x faster.
  - Use auto-encoder architecture for data compression in approximate data queries.
- **Clarkson University** — Potsdam, NY
  Research Assistant under Prof. Jeanna N. Matthews — *Sept. 2012 – Jul. 2015*
  - Build an Internet Topology Platform. The platform simulates the latency of two arbitrary Internet nodes to facilitate easy Internet structure analysis.
- **Baidu Inc.** — Shanghai, China
  System Architect — *Sept. 2011 – Aug. 2012*
  - Design and implement a customizable MySQL replication framework. This framework listens to MySQL master-slave replication messages, and rewrite the message with user-provided functions. It allows easy customization of replication behaviors.

- BigData Analysis on Hadoop. Led a 4-developer team to design and develop a Hadoop-based BigData analysis system. The system processes a daily data volume of over 10 terabytes.

- **OOCL Co. Ltd.** Shanghai, China
  Senior Software Engineer *Sept. 2005 – Sept. 2011*
  - Production Server JVM Performance Tuning. Design and implement a log analysis system for production environment JVM resource leak and memory leak tracing and tuning.
  - Design and optimization of an accounting system containing billions of rows and processing millions of queries hourly based on Oracle DBMS.

Skills

- **Computer Science:** Database, Distributed Systems, Parallel Programming, Machine Learning
- **Tech Stack:** Java, Scala, C/C++, Python, Golang, Matlab, Javascript, HTML/CSS
- **Teamworking:** Experience of leading small development teams

Education

- **The University of Chicago** Chicago, IL
  Ph.D. Computer Science *2015 – 2021*
  - Advisor: Aaron J. Elmore
- **Clarkson University** Potsdam, NY
  M.Sc. Computer Science *2012 – 2015*
  - Advisors: Jeanna N. Matthews
- **Fudan University** Shanghai, China
  B.Sc. Computer Science *2001 – 2005*

Publications

1. Chunwei Liu, **Hao Jiang**, John Paparrizos, Aaron J. Elmore *Decomposed Bounded Floats for Fast Compression and Queries VLDB 2021*
2. **Hao Jiang**, Chunwei Liu, John Paparrizos, Andrew A. Chien, Jihong Ma, Aaron J. Elmore, *Good to the last bit: Data-Driven Encoding with CodecDB*, *SIGMOD 2021*
3. **Hao Jiang**, Chunwei Liu, Qi Jin, John Paparrizos, Aaron J. Elmore, *PIDS: Attribute Decomposition for Improved Compressionand Query Performance in Columnar Storage*, *VLDB 2020*
4. Chunwei Liu, McKade Umbenhower, **Hao Jiang**, Aaron J. Elmore, *Mostly Order Preserving Dictionaries*, *ICDE 2019*
5. **Hao Jiang**, Aaron J. Elmore, *Boosting Data Filtering on Columnar Encoding with SIMD*, *DaMon 2018*

6. Dixin Tang, **Hao Jiang**, Aaron J. Elmore, *Adaptive Concurrency Control: Despite the Looking Glass, One Concurrency Control Does Not Fit All*, *CIDR 2017*

7. **Hao Jiang**, Yaoqing Liu, Jeanna N. Matthews, *IP Geolocation Estimation using Neural Networks with Stable Landmarks*, *IEEE INFOCOM Workshop GI 2016*

8. Wenjin Hu, Long Zhang, **Hao Jiang**, Jeanna N. Matthews, *A Quantitative Study of Virtual Machine Live Migration*, *CAC 2013*