Joshua A. Engels

jengels@mit.edu | 301-661-4539 | Cambridge, MA
GitHub: github.com/JoshEngels **Portfolio:** joshengels.com **Google Scholar:** [link]

EDUCATION

Rice University   **August 2018 - December 2021**
Bacheler of Science in Computer Science   Houston, TX
Bachelor of Arts in Mathematics
Graduated summa cum laude (4.00 GPA)

Massachusetts Institute of Technology   **August 2023 - Present**
PhD in Computer Science   Cambridge, MA

RESEARCH EXPERIENCE

Massachusetts Institute of Technology   **August 2023 - Present**
PhD Student, Advisor: Dr. Julian Shun   *Cambridge, MA*
- Working on the theory and practice of efficient high dimensional nearest neighbor search, clustering, and machine learning.

ThirdAI Corp   **August 2021 - May 2023**
Artificial Intelligence Engineer   *Houston, TX*
- Created DESSERT, a general algorithm for performing vector-set search with vector-set queries. When applied to the MSMarco passage retrieval task, DESSERT speeds up an existing state of the art method (ColBERT) by 5X.

Rice University Sketching and Hashing Lab   **August 2020 - August 2021**
Undergraduate Researcher, PI: Dr. Anshumali Shrivastava   *Houston, TX*
- Implemented and benchmarked FLINNG, a high performance C++ near neighbor search algorithm that uses locality sensitive hashing and group testing to achieve up to a 10x query latency speedup against SOTA methods on high dimensional genome, URL, and embedding datasets.
- Proved that FLINNG solves the nearest neighbor problem in sublinear time.

Rice University Human Computer Interaction Lab   **January 2019 - May 2020**
Undergraduate Researcher, PI: Dr. Michael Byrne   *Houston, TX*
- Built a modular framework in ACT-R and Lisp for simulating a wide variety of paper ballot voting strategies.
- Analyzed the resultant simulated voter error rates in R to identify and characterize ballots that cause systematic voting errors.

WORK EXPERIENCE

ThirdAI Corp   **August 2021 - May 2023**
Artificial Intelligence Engineer   *Houston, TX*
- Lead engineer on near neighbor search. Implemented production versions of FLASH, MACH, and DESSERT, all of which are currently being used by customers.
- Made core contributions to ThirdAI's internal machine learning engine, including work on its computation DAG (directed acyclic graph), distributed training, and extensive sparsity-based optimizations.
- Built and maintained ThirdAI's build system, continuous integration, custom RSA based licensing scheme, and automated PyPi package uploads.

Databricks   **May 2021 - August 2021**
Software Engineering Intern   *Remote*
- Designed and built a Scala service to monitor and benchmark Kafka clusters. Given a target cluster, the service produced and consumed a constant stream of messages, allowing engineers to monitor cluster availability, message durability, and end to end latency.
- Implemented a monitoring and alerting pipeline for the service and then put it into production running against all of Databricks' Kafka clusters.

Two Sigma   **May 2020 - August 2020**
Software Engineering Intern   *Remote*
- Built Baikal, a proof of concept data storage service that cached high frequency writes into a PostgreSQL database for later ingestion into Two Sigma's file-based custom time series database, while still maintaining real time read consistency (no stale reads).
- Optimized Baikal's merged read operation, raising throughput from 0.2 MB per second to 10 MB per second (a speedup of 50X).

TEACHING AND MENTORING EXPERIENCE

MIT Graduate Application Assistance Program (GAAP)   **August 2023 - Present**
Mentor   *Cambridge, MA*
- Advising and mentoring an undergraduate student from an underrepresented background who is applying to graduate school.

Rice University Computer Science Department   **January 2020 - May 2020**
Algorithmic Thinking (Comp 182) Teaching Assistant   *Houston, TX*
- Held weekly office hours, led review sessions, and graded exams and projects for Rice's algorithms and discrete mathematics introduction class.

Rice University Catalyst Eureka Program   **September 2018 - May 2019**
Mentor   *Houston, TX*
- Mentored a high school student working on a year-long research project to investigate characteristics of popular songs.
- Helped the student learn Python, write a program to scrape and analyze historical song data, and create a final presentation for a poster symposium.

SELECTED HONORS

Louis J. Walsh Scholarship in Engineering	**2020, 2021**
Top 0.18% of ~1 million solvers on ProjectEuler.net	**Present**
1st place underclassman Rice coding challenge	**2018, 2019**
National Merit Finalist Scholarship	**2018, 2019**

PUBLICATIONS

PECANN: Parallel Efficient Clustering with Graph-Based Approximate Nearest Neighbor Search
Shangdi Yu, **Joshua Engels**, Yihao Huang, Julian Shun
Under Submission at VLDB 2024

DESSERT: An Efficient Algorithm for Vector Set Search with Vector Set Queries
Joshua Engels, Benjamin Coleman, Vihan Lakshman, and Anshumali Shrivastava
Advances in Neural Information Processing Systems 2023 (to appear)

BOLT: An Automated Deep Learning Framework for Training and Deploying Large-Scale Search and Recommendation Models on Commodity CPU Hardware
Nicholas Meisburger, Vihan Lakshman, Benito Geordie, **Joshua Engels**, David Torres Ramos, Pratik Pranav, Benjamin Coleman, Benjamin Meisburger, Shubh Gupta, Yashwanth Adunukota, Siddharth Jain, Tharun Medini, Anshumali Shrivastava
ACM International Conference on Information and Knowledge Management 2023

From Research to Production: Towards Scalable and Sustainable Neural Recommendation Models on Commodity CPU Hardware
Anshumali Shrivastava, Vihan Lakshman, Tharun Medini, Nicholas Meisburger, **Joshua Engels**, David Torres Ramos, Benito Geordie, Pratik Pranav, Shubh Gupta, Yashwanth Adunukota, Siddharth Jain
ACM Conference on Recommender Systems 2023

Practical Near Neighbor Search via Group Testing.
Joshua Engels*, Benjamin Coleman*, and Anshumali Shrivastava
Advances in Neural Information Processing Systems 2021: Spotlight talk - top 3%
∗ indicates equal contribution

Missed One! How Ballot Layout and Visual Task Strategy Can Interact to Produce Voting Errors.
Joshua Engels, Xianni Wang, Michael D. Byrne
International Conference on Cognitive Modeling 2020

CONFERENCE PRESENTATIONS

How to Deal with the Volume and Velocity Associated with Hundreds of Terabytes (and Beyond) of Genomics Data
Ken Kennedy AI and Data Science Conference 2021.

Practical Near Neighbor Search via Group Testing
NeurIPS 2021: Spotlight talk.

Missed one! How ballot layout and visual task strategy can interact to produce voting errors.
International Conference on Cognitive Modeling 2020.