

Shikhar Jaiswal

PhD Student, [Department of Computer Science](#)
University of Toronto
Advisors: [Prof. Nick Koudas](#) & [Prof. Niv Dayan](#)

✉ jaiswal@cs.toronto.edu
🎓 [Google Scholar](#)
🏠 shikharj.github.io
👤 [ShikharJ](#)

EDUCATION

University of Toronto

Doctor of Philosophy, Department of Computer Science (GPA: 4.0/4.0)

September '23 - Present

Indian Institute of Technology Patna

Bachelor of Technology, Computer Science and Engineering

July '16 - May '20

RESEARCH INTERESTS

Primary: Machine Learning for Systems and Systems for Machine Learning

Others: Information Retrieval, Algorithms and Applied Mathematics

PUBLICATIONS

OOD-DiskANN: Efficient and Scalable Graph ANNS for Out-of-Distribution Queries

Shikhar Jaiswal, Ravishankar Krishnaswamy, Ankit Garg, Harsha Simhadri & Sheshansh Agrawal

Under Submission at *Neural Information Processing Systems (NeurIPS)*, 2024. [📄](#)

MinUn: Accurate ML Inference on Microcontrollers

Shikhar Jaiswal*, Rahul Kranti Kiran Goli*, Aayan Kumar, Vivek Seshadri & Rahul Sharma

In *Proceedings of the 24th ACM SIGPLAN/SIGBED International Conference on Languages, Compilers, and Tools for Embedded Systems (LCTES)*, 2023. [📄](#)

Acting Engaged: Leveraging Play Persona Archetypes for Semi-Supervised Classification of Engagement

Benjamin D. Nye*, Mark G. Core*, Shikhar Jaiswal*, Aviroop Ghosal & Daniel Auerbach

In *Proceedings of the 14th International Conference on Educational Data Mining (EDM)*, 2021. [📄](#)

(* Denotes equal contribution.)

RESEARCH EXPERIENCE

Billion-Scale Graph ANNS for Out-of-Distribution Queries [📄](#)

Advisors: [Dr. Ravishankar Krishnaswamy](#) & [Dr. Harsha Simhadri](#)

Jan '21 - October '22, *Microsoft Research*

Developed RobustVamana, a graph-based Approximate Nearest Neighbor Search (ANNS) algorithm suitable for billion-scale text-to-image embedding retrieval, and Accurate PQ (APQ), a novel product quantization algorithm for fast and accurate distance computations. The resulting system, OOD-DiskANN, offers upto 40% latency improvement over the previous state-of-the-art at billion-scale. Parts of this work are deployed across various Microsoft Bing verticals.

► Under submission at *NeurIPS '24*

Accurate ML Inference on Resource-Constrained Devices [📄](#) [📄](#) [📄](#)

Advisors: [Dr. Prateek Jain](#), [Dr. Harsha Simhadri](#) & [Dr. Rahul Sharma](#)

June '20 - Nov '21, *Microsoft Research*

Developed an efficient bit-width demotion heuristic named Haunter, and co-designed the MinUn compiler around it. MinUn generates accurate ML models under a strict RAM budget, while being agnostic of the number representation (fixed-point/floating-point/posit). When used with posits, MinUn generates TinyML models with upto 5.1× lower RAM consumption and negligible loss in accuracy/mAP scores, compared to 32-bit floating point models.

► Published as a full paper at *LCTES '23*

SMART-E: Service for Measurement and Adaptation to Real-Time Engagement [📄](#) [📄](#)

Advisor: [Dr. Benjamin D. Nye](#)

Summer '19, *USC - Institute for Creative Technologies*

Built an active-learning based algorithmic pipeline for detecting user engagement with interactive tutoring systems. Applied feature engineering to develop a set of generalizable engagement metrics, tackling cold start and automated annotation problems using easy-to-obtain play-persona data. Analysis demonstrates the utility of play-persona data gathered during professional or quality assurance testing for training useful data mining algorithms.

► Published as a full paper at *EDM '21*

SOFTWARE

DiskANN: Fast Accurate Billion-point Nearest Neighbor Search

Devvrit, Philip Adams, Andrija Antonijevic, Deng Cai, Ningyuan Chen, Magdalen Dobson, Jerry Gao, **Shikhar Jaiswal**, Rohan Kadekodi, Ravishankar Krishnaswamy, Huisheng Liu, Jigao Luo, Dax Pryce, Harsha V. Simhadri and Suhas J. Subramanya ★ >370 📄 >110

EdgeML: Machine Learning for Resource-Constrained Edge Devices

Don K. Dennis, Yash Gaurkar, Sridhar Gopinath, Sachin Goyal, Chirag Gupta, Moksh Jain, **Shikhar Jaiswal**, Ashish Kumar, Aditya Kusupati, Chris Lovett, Shishir G. Patil, Oindrila Saha and Harsha V. Simhadri
★ >1,400 📄 >360

OPEN SOURCE CONTRIBUTIONS

Implementing Essential Deep Learning Modules

Mentor: *Marcus Edel*

Google Summer of Code 2018

Deployed implementations of Generative Adversarial Networks (GAN, Deep Convolutional GAN, Wasserstein GAN and Cycle GAN) and Restricted Boltzmann Machines (RBM and Spike and Slab RBM), achieving ~1.57x single core aggregate speeds over *Sklearn's* and *Tensorflow's* implementations. ★ >4,100 📄 >1,400

Improving SymEngine's Python Wrappers and SymPy-SymEngine Integration

Mentors: *Isuru Fernando & Sumith Kulal*

Google Summer of Code 2017

Improved the overall infrastructure of *SymEngine*, an efficient, standalone C++ Computer Algebra System (CAS), and refactored its Python wrapper *SymEngine.py*. Further introduced SymEngine as an optional core for *SymPy*, and *PyDy*, a multi-body dynamics tool-kit for speeding up their backend computations by the order of ~70x. ★ >900 📄 >250

HONOURS & ACHIEVEMENTS

- Achieved *98.71%* percentile in JEE Advanced (previously IIT-JEE) 2016 among 200,000 candidates
- Achieved *99.54%* percentile in JEE Main (previously AIEEE) 2016 among 1,200,000 candidates
- Achieved *99.13%* percentile in National Entrance Screening Test (NEST) 2016 among 40,000 candidates
- Recipient of the Kishore Vaigyanik Protsahan Yojana Scholarship in 2016 (*top 1400* students out of 100,000)
- Recipient of CBSE Award for Community Service - Human Rights and Social Equality 2013