

Dhruv Kapur

+1 (734) 496-5562 | dhruvka@umich.edu | Ann Arbor, MI | <https://dhruvk19.github.io>

EDUCATION

University of Michigan, College of Engineering

Ann Arbor, MI

BSE – Computer Science with Minor in User Experience Design, GPA - 3.73 / 4.00

Dec 2025 (expected)

Relevant coursework: Data Structures & Algorithms, Parallel Programming with GPUs,

Computer Security, Web Systems, Computer Organization, Artificial Intelligence, Practical Data Science

EXPERIENCE

Student Researcher

Jan. 2024 - Present

Researcher

- Collaborating with postdoctoral researcher at University of Michigan analyzing deidentification algorithms in depth as a continuation to my NeurIPS 2023 report in differential privacy
- Observing features of the privacy-preserving techniques that result in outliers due to noise and exploring reasons for these differences in the balance between utility and privacy

Knexus Research Corporation

May. 2023 - Aug. 2023

Intern

- Explored behavior of data deidentification in the Collaborative Research Cycle NIST Program
- Evaluated outliers in diverse populations to highlight algorithms that were vulnerable to known privacy issues exemplified in other data analytics and outlined identifiable privacy concerns with outliers
- Analyzed synthetic data in an archive of over 450 deidentified data samples produced by the NIST Diverse Community Excerpts Benchmark Data using Python libraries Pandas and Numpy
- Collaborated with mentor and team to compile findings for NeurIPS 2023 Datasets and Benchmarks Track

PUBLICATIONS

Sen, A., Task, C., **Kapur, D.**, Howarth, G. S., & Bhagat, K. (2023). Diverse community data for benchmarking data privacy algorithms. *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*. <https://openreview.net/forum?id=1ODvxEwsGk>

PROJECTS

ClashBot - Clash of Clans Discord User Bot

- Developed a bot on a virtual private server (VPS) to ensure high availability and secure data storage
- Deployed bot with discord.py API to automate event scheduling with RSVP functionality for 300+ users
- Recorded commands in real-time sent by members and tracked selected options to planned events
- Organized cluttered channels by utilizing Discord Reactions by messages sent by the bot throughout the server, making event planning more synchronous and efficient with members

Distributed Search Engine

- Built a scalable search engine with separate Index and Search servers to handle concurrent requests
- Built a custom pipeline to process web documents and construct segmented inverted indices, enabling efficient parallel data processing and search functionality
- Analyzed the efficiency of various retrieval methods like text analysis (tf-idf) and link analysis (PageRank), and parallel data processing with MapReduce
- Designed a REST API that handles distributed search queries across index segments using multi-threading

Machine Learning Text Classifier

- Designed and implemented application for text classification with supervised machine learning techniques
- Trained classifier on labeled datasets extracted from large datasets of Piazza posts in CSV format, and calculated log-probability scores for accurate topic predictions
- Achieved accurate classification results on test data with an accuracy of 87% based on testing data pool of true values

CLUBS

Vice President of Operations at Tau Epsilon Kappa (Technological Professional Fraternity), MHackers-ML Team, Michigan Data Science Team

TECHNICAL SKILLS

AWS, BeautifulSoup, C/C++, CSS, CUDA, DBMS, Flask, Git, Hadoop, HTML, JavaScript, Julia, Node.js, PyTorch, Python, R, React Native, React.js, SQL, SQLite, Swift/SwiftUI, Visual Basic, Web Scraping