SWAGNIK ROYCHOUDHURY

CS + DS Double Major ~ NLP Researcher

NYC, New York
sr6474@nyu.edu
swagnik.org
GitHub
LinkedIn
Google Scholar

SUMMARY

I am currently double majoring in Computer Science and Data Science at New York University. I'm a fully funded researcher at the NYU Ensure Lab, and hold research positions at NJIT and MSU. My work is primarily focused on natural language processing and computer vision. I am also a fully trained Indian Classical musician and state-level chess player.

SKILLS

Languages: C++, C#, Python, SQL, NoSQL, Java, Swift, Dart/Flutter, JavaScript

Technologies: AWS, CUDA, PyTorch, HuggingFace, Linux, Git, PostgreSQL, MySQL, Unity, Solidworks, Multisim, Agile + Confluence, ROS, XCode, Android Studio, Spark, React/Node

EDUCATION

2022 - 2025 **New York University, College of Art and Sciences @ Courant and CDS** **University**
Double Majoring in Computer Science and Data Science with a Mathematics Minor. 3.9 GPA. NYU Robomasters Robotics Competition Team Lead, NYU Chess Club. Fully Funded Researcher at NYU Ensure Group and Data Science Fellow at NYU Marron Institute.

2018 - 2022 **Middlesex County Academy for Science, Mathematics, and Engineering Technologies** **High School**
Electrical and Computer Engineering Concentration. Member of the National Honor Society and Technology Student Association.

2017 - 2021 **Sarbabharatiya Sangeet O Sanskriti Parishad** **University**
Pursued my Visharad Degree (B.A Equivalent) in Indian Classical Music, with multiple performances in the United States and abroad.

PUBLICATIONS

6/2024 – 9/2024 **The DISCERN Approach for Intelligent and Efficient Discernment of Robotic Task Contexts** **MSU**

- Authored paper that introduces DISCERN (Detection Image System with Commonsense Efficient Ranking Network). DISCERN is an end-to-end robotics task execution pipeline, that simplifies the task ordering step by using commonsense linguistic networks and vision language models for fast classification. Accepted to the MIT URTC 2024 Conference.

12/2022 – 2/2023 **Applications of BadNets in Spam Filters** **NYU Ensure Lab**

- Authored paper exploring applications of BadNets and backdoored models and their consequences beyond Image Recognition in the domain of natural language processing, such as Spam Filter Detection. Accepted at ICDE 23 Astride workshop.

3/2022 – 3/2023 S^2 **- Information-Theoretically Secure and Highly Efficient Search and Row Retrieval** **NJIT**

- Co-Authored paper that focuses on creating homomorphic encryption algorithms to store data securely and more efficiently than current state-of-the-art systems. Responsible for developing a suite of eighty programs to test, modify, and provide test results for the algorithms. They were implemented in an AWS EC2 environment. Accepted at the VLDB 23 conference.

AWARDS

12/2023 **Goldwater Scholarship Nominee** **NYU**

- One of four students nominated by NYU for the Barry Goldwater Scholarship, one of the most prestigious undergraduate national scholarships. Currently going through the final selection stage.

2023 - 2024 **3x DURF Grant Recipient** **NYU**

- Awarded two research grants by NYU CAS's Dean for my work in fairness and biases in language models, and my work in developing NLP architectures for Indian Classical Music. Awarded one conference grant to present my research related to DISCERN at MIT.

2022-2024 **NYU - Dean's List** **NYU**

- Awarded Dean's List for the 2022-2023 year and 2023-2024 for exemplary academic achievement.

EXPERIENCE

9/2024 – Current **Brain-MRI Researcher** **NYU CDS**

- Developing masked autoencoders and vision transformers to predict downstream neuropsychiatric symptoms of Alzheimer's. The deep-learning models are trained on 3D MRI Scans and Clinical Survey Data.

6/2024 – Current **Human-Robotics Collaboration Researcher** **Montclair State University**

- Developing the DISCERN approach, which integrates common sense knowledge (CSK), linguistic networks, and vision language model for robotics in a human-robot collaboration environment. This approach doesn't require specialized hardware or training - for example, it is compatible with nearly 20,000 of the classes of ImageNet-21k without any pertaining. This also allows the model to generalize task execution ordering easily without any training and minimal human intervention.

6/2024 **LLM Consultant for attend.ai** **attend.ai**

- Implemented LLM integration and deployment in preparation for attend.ai's third YC interview. Created user profile + customization system via chromaDB for a personalized chatbot.

1/2024 – 3/2024 **Data Science Intern** **NYU Marron Institute**

- Developing visualizations to create info-graphics for de-notified tribes in India, particularly in explaining and raising awareness for labor exploitation and human trafficking in suburban areas.

10/2023 – 2/2024 **Data Science Student Fellow** **NYU Marron Institute**

- Working on data collected by the Marron Institute regarding runaway and homeless youth in NYC in an attempt to disrupt the human trafficking industry. The goal of this project is to identify support services and resources to best aid youth at risk.

9/2023 – 1/2024 S^2 **- Demonstration Paper** **Paper, NJIT**

- Working on a demonstration paper for a prior paper we submitted to VLDB 2023. We are working on generalizing the encryption system to any database with variable numbers of column and rows, as well as supporting string and date data types.

8/2023 – 2/2024 **ICMLM, A Language Model for Indian Classical Music** **Paper, NYU Ensure Lab (Poster)**

- Focusing on using SOTA transformer architectures for generating Indian classical music using a handmade, first-of-its-kind, dataset. Additionally, we compare the model's performance against in-context learning and fine tuning with GPT4, Claude 2, and LLAMA 2. Poster presented at New York University's Undergraduate Research Conference 2024.

7/2023 – 3/2024 **Fairness and Bias Issues in Large Language Models** **Paper, NYU Ensure Lab**

- Using datasets with sensitive attributes (ie race, gender, age) we are testing various language models such as GPT, Claude, LLAMA, and Bard to tease out inherent biases that these language models may have.

6/2023 – 11/2023 **NSF REU Research Internship at University of California, Irvine** **UCI**

- Developing a powerful visualizer for databases that include spatial and temporal data. The tool is able to use wifi connectivity data to precisely estimate the occupancy of rooms, floors, and buildings within a campus, helpful for first responders when trying to evacuate a building during an emergency. I am currently working on expanding this tool to aid in wildfire visualization via interpolation of drone images of the fires. Poster Presented at UC Irvine's undergraduate research symposium.

9/2022 – Current **Competition Team Lead @ NYU Robomasters** **NYU**

- Manage a team of 11 members as Competition Team Lead at NYU's Robomasters Robotics team. Responsibilities include developing computer simulations of the competition, developing CV algorithms for our autonomous robots, working on CAD for the robots, training drivers for the competition, and smoke-testing embedded functionalities of our robots. Working with Vision Transformers and ROS.

6/2022 – 1/2023 **Software Developer & Data Science Intern** **INVIDI Technologies**

- Worked on ETL (extract, transform, load) of advertisement impression data that INVIDI collects from its clients in India. Using AWS Redshift, Sagemaker, and S3, I developed RNNs for time series analysis to derive actionable insights from the data.

1/2021 – 10/2022 **Kathak Saangi** **iOS App Store Link | Google Play Store Link**

- Creator of iOS/Android app Kathak Saangi, a companion app for Kathak Dancers. Available internationally, with over 10,000 downloads. The iOS version was developed in XCode with Swift, and the Android version was developed in Android Studio with Dart/Flutter.