

HARSHIT JOSHI New Delhi, India

(+91) 991 026 8058 | joshharshit@gmail.com | github.com/duskybomb | linkedin.com/in/harshitjos | harshitjoshi.in

EDUCATION

Cluster Innovation Centre, University of Delhi
Bachelor of Technology (B. Tech)
Information Technology and Mathematical Innovations

July 2017 - August 2021
Overall Percentage: 87.88%
Department Rank: 2

INTERESTS

NLP, Program Synthesis, AI4SE, Computational Social Science

RESEARCH EXPERIENCE

Microsoft
Research Fellow: PROSE

November 2021 - Present
India - Remote

- Improving experiences in spreadsheet-based environments. Predominantly working on program repair for excel formulae.

Multimodal Digital Media Analysis Lab (MIDAS@IIITD)
Research Intern

May 2020 - April 2021
New Delhi, India

- Built Deep Learning NLP models for social good, focusing on contextualizing users posts on social media (Twitter, Reddit)
- Proposed interpretable models for suicide ideation detection and analyzed emotional changes of users over time.

Defence Research and Development Organisation (DRDO)
Research Intern

June 2019 - October 2019
New Delhi, India

- Used CityScape Dataset for Image Segmentation through the PyTorch implementation of DeepLabV3+
- Fine-tuned the Image Segmentation model for two classes: void space and obstacle for the cognitive navigation and mapper.

PUBLICATIONS

* Co-First Authors

H. Joshi, J. Cambronero, S. Gulwani, V. Le, I. Radicek and G. Verbruggen. *Repair Is Nearly Generation: Multilingual Program Repair with LLMs*. Under Review 2022.

R. Bhavishi*, **H. Joshi***, J. Cambronero, A. Fariha, S. Gulwani, V. Le, I. Radicek and Ashish Tiwari. *Neurosymbolic Repair for Low-Code Formula Languages*. In In Proceedings of the ACM on Programming Languages (OOPSLA) 2022.

R. Sawhney*, **H. Joshi***, A. Nobles*, and R. R. Shah. *Towards Emotion-and Time-Aware Classification of Tweets to Assist Human Moderation for Suicide Prevention*. In International AAAI Conference on Web and Social Media 2021.

R. Sawhney*, **H. Joshi***, R. R. Shah, and L.Flek. *Suicide Ideation Detection via Social and Temporal User Representations using Hyperbolic Learning*. In North American Chapter of the Association for Computational Linguistics 2021.

R. Sawhney*, **H. Joshi***, L.Flek, and R. R. Shah. *Phase: Learning Emotional Phase-Aware Representations for Suicide Ideation Detection on Social Media*. In European Chapter of the Association for Computational Linguistics 2021.

R. Sawhney, **H. Joshi**, S. Gandhi, D. Jin, and R. R. Shah. *Robust Suicide Risk Assessment on Social Media via Deep Adversarial Learning*. In Journal of the American Medical Informatics Association 2021.

R. Sawhney, **H. Joshi**, S. Gandhi, and R. R. Shah. *Towards Ordinal Suicide Ideation Detection on Social Media*. In ACM International Conference on Web Search and Data Mining 2021.

R. Sawhney, **H. Joshi**, S. Gandhi, and R. R. Shah. *A Time-aware Transformer based Model for Suicide Ideation Detection on Social Media*. In Conference on Empirical Methods in Natural Language Processing 2020.

PROFESSIONAL EXPERIENCE

Supedio GmbH
Research Software Engineer

August 2021 - October 2021
Germany - Remote

- Worked on biomedical product entity linking for matching hospital's clinical orders in the supplier's product catalog.
- Built automation tools for the sales team, reducing person-hours employed per day by 400%.

Supedio GmbH
Software Engineer Intern

January 2021 - June 2021
Germany - Remote

- Engineered heuristics for extracting financial data from tables of digital invoices, orders, delivery notes.
- Built algorithmic pipelines for address deduplication, finding 18% duplicates in client's "processed" database.

Cronycle Ltd.

Software Engineer Intern: Data Science

January 2019 - July 2019

U.K. - Remote

- Migrated batch jobs for retrieving RSS to live at production scale using Kafka and ElasticSearch, reducing latency by 5 min.
- Increased RSS collection dump by 10% by identifying new data sources and processing them to MongoDB.

TECHNICAL STRENGTHS

Computer Languages

Proficiency: Python, SQL, **Familiar:** JAVA, R, C/C++, Go

Software & Tools

PyTorch, MongoDB, Git, Elastic Search, PostgreSQL, Flask, Airflow, Kafka

NOTABLE PROJECTS

Google Summer of Code 2018: invoice2data

Student Developer at Debian Project

April 2018 - August 2018

Remote

- Improved Tesseract-OCR integration and increased code coverage by 16% by adding tests for functions.
- Developed a GUI application for three major OS (Linux, macOS, Windows) using Python and PyQt and setting up CI/CD.

Content Based Video Retrieval System

Semester Long Project

Prof. Dr. Shobha Bagai

January 2019 - Present

- Divided movie trailers into different scenes using PySceneDetect and evaluated image captioning models on still frames.
- Implemented Deep Learning models to retrieve video clips from natural language on TaCoS dataset [Link]

Cognitive Mapping and Planner for Navigation

Defence Research and Development Organisation

Scientist. S.P. Mishra

June 2019 - October 2019

- Worked on the navigation of bot in the Gazebo environment.
- Used CityScape Dataset for Image Segmentation and used the features in the Gazebo environment
- Built an algorithm for depth perception using image segmentation results and creating disparity maps.

Strategizing Fantasy Football

Probability and Statistics

Dr. Sonam Singh

March 2019 - April 2019

- Created new visualizations and performance metrics for better analysis of players' points per game output.
- Used Gradient Boosting Trees, Stochastic Gradient Regressor, and Linear Regression to regress points scored against 32 features engineered through priori.

POSITION OF RESPONSIBILITY

Govt. of India's Institutions Innovation Council

Student Coordinator

November 2018 - October 2019

University of Delhi

- We achieved a 4-star rating, given by MHRD, Govt. of India
- Lead a team of top 10 innovators, who oversee the innovations across Delhi University.

HashInclude - Computer Science Society

External Affairs

August 2018 - August 2019

Cluster Innovation Centre

- Spearheaded a team of 40 people to conduct professional shows, exhibitions and talks successfully
- Conducted workshops on Open Source Development and Linux 101

AWARDS AND ACHIEVEMENTS

Received Honorable Mention at 2020 **COMAP's** Mathematical Contest in Modeling (MCM). **Only team from India** to get Honorable Mention.

Invited talk at **PyData Delhi Conference 19** on "Quantitative Finance with R"

Founding member of football team, Cluster Innovation Centre.

Mathematical Finance Scholar under Focus Areas in Science and Technology **Summer Fellowship 2019**

Google Summer of Code 2018 with Debian Project

Qualified for **ACM-ICPC 2018** Kolkata Kanpur Site contest held at UIET Kanpur

VOLUNTARY SERVICES

Member of Programme Committee CA2MH workshop at ICML 2021

Student Volunteer at EMNLP 2020, EACL 2021, ICWSM 2021