HARSHIT **JOSHI**

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

EDUCATION

Cluster Innovation Centre, University of Delhi *July 2017 - August 2021*
Bachelor of Technology (B. Tech) Overall Percentage: 87.88%
Information Technology and Mathematical Innovations Department Rank: 2

INTERESTS

Program Synthesis, Machine Learning for Programming Languages and Software Engineering, NLP

RESEARCH EXPERIENCE

Microsoft Research Fellow: PROSE, Redmond, US November 2021 - Present (Bengaluru, India - Remote)

- Developed a small language model for spreadsheets (60M) which is competent and even outperforms much larger LLMs (175B) in formula repair and formula autocompletion. This work was covered by media outlets.
- Built SOTA production-ready neurosymbolic program repair framework for Excel and PowerApps.
- Created multilingual repair framework, RING, built on top of Large Language Models and leverages compiler diagnostics.

Multimodal Digital Media Analysis Lab 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.

DRDO, Govt. of India 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, A. Ebenezer, J. Cambronero, S. Gulwani, A Kanade, V. Le, I. Radicek and G. Verbruggen. *FLAME: A small language model for spreadsheet formulas* Under Submission, 2023 `https://arxiv.org/abs/2301.13779`.

H. Joshi, J. Cambronero, S. Gulwani, V. Le, I. Radicek and G. Verbruggen. *Repair Is Nearly Generation: Multilingual Program Repair with LLMs.* Accepted at AAAI 2023 `https://arxiv.org/abs/2208.11640`.

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, C# **Familiar**: JAVA, R, C/C++, Go, F#
Software & Tools	PyTorch, MongoDB, Git, Elastic Search, PostgreSQL, Flask, Airflow, Kafka

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

Honourable Mention at **ACM-ICPC 2018** Regionals (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