

# Mukul Singh

RESEARCHER · MICROSOFT RESEARCH

✉ +1 (206)-407 8243 | 📩 mukulsingh105@gmail.com | 🏷 Homepage | 📱 mukulsingh105 | 💬 mukulsingh105

## Achievements

2025	<b>Dagstuhl Invited Guest</b>	Invited for the Automated Programming and Program Repair Seminar at Dagstuhl	Dagstuhl, Germany
2025	<b>ICSE Keynote Speaker</b>	Keynote demo at AIWARE conference co-hosted with ICSE 2025	Seoul, South Korea
2024	<b>ICSE Best Paper Award</b>	Awarded at International Conference for Software Engineering 2024	ICSE, USA
2024	<b>SMU Invited Talk</b>	Invited for a research keynote at Singapore Management University (SMU)	SMU, Singapore
2023	<b>NeurIPS Best Reviewer Award</b>	Awarded at NeurIPS Table Representation Learning Workshop 2023	NeurIPS, USA
2023	<b>VLDB Tabular Data Analysis Invited Speaker</b>	Presented my work on Spreadsheet Formatting [Video]	VLDB, Canada
2022	<b>Sustainability Research and Leadership Award</b>	Awarded for my work on AI for vehicle automation	Tata Power
2021	<b>Young Researcher</b>	Awarded for my Bachelor thesis by NSS and Govt. of India	Govt. of UP
2020	<b>TIDE Fellowship</b>	Awarded Tech Entrepreneurship Award and Incubation by MeitY, Govt. of India	MeitY, Govt. of India
2018	<b>Merit Scholarship</b>	Awarded full merit scholarship based on scholastic performance by Govt. of India	IIT Delhi
2017	<b>KVPY</b>	Fellowship under <b>Kishore Vaigyanik Protsahan Yojana</b> by IISc Bangalore and Govt. of India	IISc, Bangalore
2016	<b>InPhO</b>	Awarded Merit Certificate for outstanding performance in International Physics Olympiad	IAPT, Delhi

## Industrial Experience

### Microsoft Research (PROSE Group)

RESEARCHER

Redmond, USA

Jul 2022 - Present

#### Code Generation

- Working on improving code generation using large language models through (1) inference based (2) agentic and (3) fine tuned frameworks.
- My research powers features in GitHub Copilot, Excel Copilot, Visual Studio and Power Query with over 1B monthly users.

#### Diffusion

- Working on developing novel combined transformer and diffusion architectures for structured and sequential generation (text, code, language)
- The initial papers I published in this area introduced these new frameworks and are foundational to diffusion for structured outputs.

#### Table Understanding

- Designed a novel multi-modal tool for generating table formatting rules in Excel from natural language and user examples
- The technology is being integrated in Excel to suggest formatting rules to users and is estimated to impact over 1 Billion users

#### UI and Action Research

- Designed a novel transformer model over natural language queries along with on-screen input to automate UI based workflows
- The system is being tested for integration in Windows and Outlook for action recommendation and was recently presented at IJCAI

### Microsoft Research

RESEARCH INTERN

Redmond, USA (Remote)

Jun 2021 - May 2022

- Developed a neurosymbolic AI system for automatic spreadsheet formatting and visualizations using user examples and data context.
- The work led to multiple papers in VLDB and SIGMOD seeding the research direction of inductive learning for table formatting.

### Meta (Facebook)

AI RESEARCH INTERN

London, UK

May 2022 - July 2022

- Curated and designed an Auto-encoder based anomaly detection system to find manufacturing defects in components using assembly images

### American Express

AI RESEARCH INTERN

Gurgaon, India

May 2021 - Jul 2021

- Developed a Deep Learning based tool to predict open rates of Email subject lines and give recommendations to improve their performance

### Robert Bosch GmbH

RESEARCH INTERN

Bangalore, India

Dec 2019 - June 2020

- Researched and developed vehicle reaction time estimation model using driving pattern data to evaluate vehicle chassis health

## Education

### Indian Institute of Technology Delhi

Minor Specialization in Artificial Intelligence

CGPA - 9.6/10.0

Delhi, India

2021 - 2022

### Indian Institute of Technology Delhi

B.Tech in Computer Science and Engineering

CGPA - 9.2/10.0

Delhi, India

2018 - 2022

# Academic Experience

---

## Center for Automotive Research and Tribology (CART)

RESEARCH ASSISTANT

IIT Delhi

December 2020 - June 2022

- CART is a research and innovation lab at IIT Delhi focused on inter-disciplinary research in the area of electric vehicle and power storage
- Working on multiple projects involving applications of cybersecurity, machine learning and blockchain in electric vehicles and power systems

## Department of Computer Science and Engineering

IIT Delhi

TEACHING ASSISTANT

June 2021 - June 2022

- Introduction to Computer Science (COL100): Conducted tutorials and help sessions, and designed testbench for assignments
- Computer Architecture (COL216): Designed assignments and conducted labs based on FPGA and advised students for final project

## Selected Publications

---

### • RAR: Retrieval Augmented Retrieval for Code Generation in Low Resource Languages.

Mukul Singh, Avik Dutta, Vu Le, Sumit Gulwani, Gust Verbruggen

Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), Nov 2024, Miami, USA

### • One-to-many testing for code generation from (just) natural language.

Mukul Singh, Mansi Uniyal, Vu Le, Sumit Gulwani, Gust Verbruggen

Conference on Empirical Methods in Natural Language Processing (EMNLP 2024), Nov 2024, Miami, USA

### • Diffusion is a code repair operator and generator.

Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen

The Conference and Workshop on Neural Information Processing Systems (NeurIPS 2025), Dec 2025, San Diego, USA

### • Tabularis Revilio: Converting Text to Tables.

Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen

33rd ACM International Conference on Information and Knowledge Management (CIKM), Oct 2024, Boise, USA

### • Learning Email Folder Classification Rules by Demonstration.

Mukul Singh, Jose Cambronero, Vu Le, Sumit Gulwani, Gust Verbruggen

38th AAAI conference (AAAI), Jan 2024, Vancouver, Canada

### • DATAVINCI: Learning to Repair String Data.

Mukul Singh, Jose Cambronero, Vu Le, Sumit Gulwani, Gust Verbruggen

23rd AAAI conference (AAAI), Jan 2023, Washington, USA

### • CODEFUSION: A Pre-trained Diffusion Model for Code Generation.

Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Carina Negreanu

Conference on Empirical Methods in Natural Language Processing (EMNLP 2023), Dec 2023, Singapore

### • CORNET: Learning Formatting Rules for Tabular Data by Example.

Mukul Singh, Jose Cambronero, Carina Negreanu, Vu Le, Sumit Gulwani, Gust Verbruggen

49th International Conference on Very Large Data Bases (VLDB), August 2023, Vancouver, Canada

### • Example based Automated Spreadsheet Formatting Rule Generation.

Mukul Singh, Jose Cambronero, Carina Negreanu, Vu Le, Sumit Gulwani, Gust Verbruggen

49th International Conference on Very Large Data Bases (VLDB) – Demonstration, August 2023, Vancouver, Canada

### • FORMAT5: Multi-modal Table Formatting with Natural Language and Examples.

Mukul Singh, José Cambronero, Sumit Gulwani, Vu Le, Carina Negreanu, Elnaz Nouri

International Conference on Machine Learning (ICML), July 2023, Hawaii, USA

### • Personalized UI suggestions for low-code automation platforms.

Saksham Gupta, Mukul Singh, Gust Verbruggen, Sumit Gulwani, Vu Le

International Conference on Software Engineering (ICSE), January 2023

### • EMFORE: Online Learning of Email Folder Classification Rules.

Mukul Singh, Gust Verbruggen, José Cambronero, Sumit Gulwani and Vu Le

International Joint Conference on Artificial Intelligence (IJCAI), May 2023, Macao, China

### • Learning from mistakes: Example based personalized online learner.

Mukul Singh, Gust Verbruggen, José Cambronero, Sumit Gulwani and Vu Le

24th AAAI conference (AAAI), Jan 2024, Vancouver, Canada

### • Deep Learning Model Based CO2 Emissions Prediction using Vehicle Telematics Data.

Mukul Singh, Rahul Dubey

IEEE Transactions on Intelligent Transport Systems (IEEE-ITS), August 2021

### • Transfer Learning based Ensemble SVM for Automated COVID-19 Detection using lung CT Scan Image Data.

Mukul Singh, Shrey Bansal, Bijaya Ketan Panigrahi, Rahul Dubey

Springer Medical and Biological Engineering and Computing (MBEC), October 2020

### • Multi-Objective Genetic Algorithm Based Deep Learning Model for Automated COVID-19 Detection.

Mukul Singh, Shrey Bansal, Bijaya Ketan Panigrahi, Rahul Dubey

Springer Journal of Medical and Biological Engineering (JMBC), August 2020

### • An RNN-LSTM based Remaining Battery Capacity Prediction Model.

Mukul Singh, Shrey Bansal, Bijaya Ketan Panigrahi, Akhil Garg

ASME Journal of Computing and Information Science in Engineering (JCISE), January 2021

- **Vehicle Telematics: Application of Machine Learning and Big Data at Edge.**  
Mukul Singh, Rahul Dubey  
*Elsevier Artificial Intelligence and Machine Learning for Edge Computing*, March 2021
- **Execution-guided within-prompt search for programming-by-example.**  
Gust Verbruggen, Mukul Singh, Ashish Tiwari, Vu Le, Sumit Gulwani  
*International Conference on Language Representation*, April 2025, Singapore
- **TeCoFeS: Text Column Featurization using Semantic Analysis.**  
Ananya Singha, Mukul Singh, Ashish Tiwari, Vu Le, Sumit Gulwani  
*Nations of the Americas Chapter of the ACL (NAACL)*, April 2025, New Mexico, USA
- **From Words to Code: Harnessing Data for Program Synthesis from Natural Language.**  
Anirudh Khatry, Mukul Singh, Joyce Cahoon, Jordan Henkel, Shaleen Deep, Avrilia Floratou, Sumit Gulwani, Vu Le, Ashish Tiwari  
*International Conference on Very Large Databases (VLDB)*, March 2023
- **INSTRUCTEXCEL: A Benchmark for Natural Language Instruction in Excel.**  
Justin Payan, Mukul Singh, Swaroop Mishra, Carina Suzana Negreanu, Christian Poelitz, Chitta Baral, Subhro Roy, Elnaz Nouri  
*Conference on Empirical Methods in Natural Language Processing (EMNLP)*, June 2023
- **TST<sup>R</sup>: Target Similarity Tuning Meets the Real World.**  
Anirudh Khatry, Mukul Singh, Sumit Gulwani, Vu Le, Gust Verbruggen  
*Conference on Empirical Methods in Natural Language Processing (EMNLP)*, June 2023
- **COOPER: Learning what to teach models for code generation.**  
Anirudh Khatry, Mukul Singh, Sumit Gulwani, Vu Le, Gust Verbruggen  
*Association for Computational Linguistics (ACL)*, August 2023
- **An empirical study of validating synthetic data for formula generation.**  
Usneek Singh, Mukul Singh, Sumit Gulwani, Vu Le, Gust Verbruggen  
*Association for Computational Linguistics (ACL)*, August 2023
- **Semantically Aligned Question and Code Generation for Automated Insight Generation.**  
Ananya Singha, Bhavya Chopra, Anirudh Khatry, Sumit Gulwani, Austin Henley, Vu Le, Chris Parnin, Mukul Singh, Gust Verbruggen  
*International Conference on Software Engineering (ICSE)*, August 2024, USA

## Patents

---

- **Execution Guided Within-Prompt Search With Large Language Models.**  
Mukul Singh, Ashish Tiwari, Vu Le, Sumit Gulwani, Gust Verbruggen
- **System and Method for Intelligent Chart Design Recommendations in Spreadsheet Applications Using Large Language Models.**  
Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Multistep retrieval for grammar based code generation.**  
Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Dynamic Refinement of Langauge Model Evaluation Checklists.**  
Mukul Singh, Mansi uniyal, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Tabular Data Augmentation Through Semantic Text Analysis and Labeling.**  
Mukul Singh, Ananya Singha, Vu Le, Sumit Gulwani
- **Recognizing tables in unstructured text.**  
Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Vision chain-of-thought reasoning for multimodal language models.**  
Mukul Singh, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Code generation system using pre-trained diffusion model.**  
Mukul Singh, Jose Cambronero, Carina Negreanu, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Data based rule generation for tabular task automation suggestion and ranking.**  
Mukul Singh, Jose Cambronero, Carina Negreanu, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Learning Syntactic and Semantic String Repairs.**  
Mukul Singh, Jose Cambronero, Carina Negreanu, Vu Le, Sumit Gulwani, Gust Verbruggen
- **Target Property Selection Techniques for Learning What to Teach Language Models for Code Generation.**  
Anirudh Khatry, Mukul Singh, Gust Verbruggen, Sumit Gulwani, Vu Le
- **Automated recommendation and ranking of personalized user interface actions in task and data automation platforms.**  
Saksham Gupta, Mukul Singh, Gust Verbruggen, Sumit Gulwani, Vu Le
- **Automated system for battery health and capacity prediction using charge and voltage statistics.**  
Mukul Singh, Bijaya Ketan Panigrahi, Akhil Garg

## Service

---

- **Senior Area Chair** for EMNLP 2025, ACL 2025 and NAACL 2025 Sept 2025
- **Area Chair** for EMNLP 2024 (Ethics Chair), ICSE 2024 (Ethics Chair), CIKM 2024 and ACL 2024 (Dec 2024)
- **Program Committee** for ACL 2023, EMNLP 2023 & 2024, AAAI 2024, NeurIPS (Aug 2023 - Present)
- **Reviewer** for IEEE TIV, IEEE ITS, Springer Nature MBEC and JMBE (Nov 2021 - Present), SIGMOD, OOPSLA (Jan 2024 - Present)
- **Research Advisor** for CART, IIT Delhi (Feb 2022 - Present)