

Rupak Sarkar

8601 34th Avenue
College Park, MD-20740
☎ +1(631)652-5197
✉ rupak@umd.edu
📄 styx97.github.io

Education

- 2021-present **Ph.D. Student**, Department of Computer Science, University of Maryland, College Park, Maryland, USA.
Advisor: Prof. Philip Resnik
- 2016-2020 **B. Tech**, Kalyani Government Engineering College, Kalyani, India.
Discipline: Computer Science and Engineering, CGPA: 8.98/10.00

Experience

Research Scientist Intern at Microsoft Reserach, Redmond, Washington, USA.

- June 2024 - Worked on understanding the information needs of users in human-AI conversations, and
Aug 2024 how we can make conversational agents better serve a diverse range of human queries.
Manuscript in Progress

Research Scientist Intern at Adobe, College Park/Hybrid, Maryland, USA.

- June 2022 - Designed and Implemented a framework for identifying related concepts in an educational
Aug 2022 document and fetching intuitive explanations of concepts from the web. Worked with Prof. Ani Nenkova and Dr. Varun Manjunatha.

Research Intern at Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (remote).

- Sep 2020 - Course Research Engineer for 11-865/11-665: *Tracking Political Sentiments Using Machine*
Dec 2020 *Learning* with instructors Dr. Ashiqur R. Khudabukhsh, Prof. Mark Kamlet and Prof. Tom Mitchell.

Publications

- 2025 “*Understanding Common Ground Misalignment in Goal-Oriented Dialog: A Case-Study with Ubuntu Chat Logs*”, by Rupak Sarkar, Neha Srikanth, Taylor Hudson, Rachel Rudinger, Claire M. Bonial and Philip Resnik. [**Under Review at ARR**]
- 2024 “*PairScale: Analyzing Attitude Change in Online Communities*”, by Rupak Sarkar, Patrick Y Wu, Kristina Miler, Alexander Hoyle* and Philip Resnik*. [**NAACL 2025 Findings (long paper)**]
- 2024 “*Pregnant Questions: The Importance of Pragmatic Awareness in Maternal Health Question Answering*”, by Neha Srikanth*, Rupak Sarkar*, Rachel Rudinger and Jordan Boyd-Graber. [**NAACL 2024 Main (long paper)**]
- 2023 “*Making the Implicit Explicit: Implicit Content as a First Class Citizen in NLP*”, by Alexander M. Hoyle*, Rupak Sarkar*, Pranav Goel and Philip Resnik. [**EMNLP 2023 Main (long paper)**]
- 2022 “*Are Neural Topic Models Broken?*”, by Alexander M. Holye, Pranav Goel, Rupak Sarkar, and Philip Resnik. [**Findings of EMNLP 2022 (long paper)**]
- 2022 “*Fringe News Networks: Dynamics of US News Viewership following the 2020 Presidential Election*”, by Ashiqur R. KhudaBukhsh*, Rupak Sarkar*, Mark S. Kamlet, and Tom M. Mitchell. [**14th ACM Web Science Conference 2022, 269-278**]

- 2021 “Are Chess Discussions Racist? An Adversarial Hate Speech Data Set”, by Rupak Sarkar and Ashiqur R. KhudaBukhsh. [AAAI 2021, Best AAAI-21 Student Abstract 3-Minute Presentation (Student Abstract)]
- 2021 “We Don’t Speak the Same Language: Interpreting Polarization Through Machine Translation”, by Ashiqur R. KhudaBukhsh*, Rupak Sarkar*, Mark S. Kamlet, and Tom M. Mitchell. [AAAI 2021, (long paper) acceptance rate : 21%]
- 2020 “Social Media Attributions in the Context of Water Crisis”, by Rupak Sarkar*, Sayantan Mahinder*, Hirak Sarkar, and Ashiqur R. KhudaBukhsh. [EMNLP 2020 Main, (long paper) acceptance rate : 22.4%]

Talks

- Feb 2021 *Are Chess Discussions Racist? An Adversarial Hate Speech Data Set*, Winner of Best AAAI-21 Student Abstract 3-Minute Presentation, AAAI 2021 (SA-21).
- Nov 2020 *The Non-native Speaker Aspect: Indian English in Social Media*, W-NUT, EMNLP 2020.
- Nov 2020 *Social Media Attributions in the Context of Water Crisis*, EMNLP 2020.

Selected Media Coverage

- NYT “For Trump’s Backers in Congress, w ‘Devil Terms’ Help Rally Voters”, *The New York Times*, October, 2022
- CMU Press “The Science of Political Polarisation”, *CMU Science Magazine (cover story)*, August, 2021
- BBC “India Covid-19: AI Shows Pakistani Twitter Prayed for Neighbour”, *BBC*, July, 2021
- WIRED “Why a YouTube Chat About Chess Got Flagged for Hate Speech”, *WIRED*, March, 2021
- CMU Press “AI May Mistake Chess Discussions as Racist Talk”, *CMU Press Release*, February, 2021
- Independent “Newsmax and OAN: How are the ultra-conservative cable channels coping without Trump in the White House?”, *Independent*, February, 2021
- Houston Chronicle “Don’t blame Fox News for the attack on the Capitol”, January, 2021 (Houston Chronicle, The Conversation)
- WIRED “The Left and the Right Speak Different Languages — Literally”, *WIRED*, October, 2020
- CMU Press “Even our language is polarized”, *CMU Press Release*, October, 2020
- Futurity “Mask vs. Muzzle: Even Words Are Now Polarized.”, *Futurity*, October, 2020
- Anandabazar (Bengali) “Two scientists researched about the polarization of politics that has crept into the language”, *Anandabazar Patrika*, October, 2020

Awards

- 2021 Was awarded the **Dean’s Fellowship - PhD, CS** by University of Maryland, College Park.
- 2021 Won the **Best AAAI-21 Student Abstract 3-Minute Presentation** award for “Are Chess Discussions Racist? An Adversarial Hate Speech Data Set” at AAAI-2021 out of 400+ submissions.
- 2019 Won the **Smart India Hackathon 2019 Software Edition** with a prize money of 75,000 INR as part of the team representing our college. Our problem statement (AM1) was focused on building an Augmented Reality App for the popular e-furniture company, PepperFry.

Skills and Interests

- Tools** Python (PyTorch, keras), R, C++, MATLAB, SQL
- Research Interests** NLP, Computational Social Science, Document Representations, Polarisation in the Web

Other Projects

Research Internship at IIT Kharagpur, Kharagpur, West Bengal, India.

July 2019 - Designed and implemented a framework for detecting Deep Fake videos in Social media
Mar 2020 using facial landmark distances for my summer internship project and subsequent final year B.Tech dissertation, supervised by Prof. Rajat Subhra Chakraborty.

Data Science Intern at *dexecure.com*.

Jan 2017 - Used Google BigQuery to find meaningful performance insights from the Chrome User
Feb 2017 Experience Report published by Google, findings of which were featured in this *blog post*.
Worked as a remote freelancer with *Dexecure*, a startup based in Singapore.

References

Prof. Philip Resnik, Professor, Institute for Advanced Computer Studies and Department of Linguistics, University of Maryland, College Park, MD, USA.
resnik@umd.edu

Prof. Mark S. Kamlet, University Professor of Economics and Public Policy and Provost Emeritus, Carnegie Mellon University, PA, USA.
kamlet@cmu.edu

Dr. Ashiqur R. KhudaBukhsh, Assistant Professor, Golisano School of Computing and Information Sciences, Rochester Institute of Technology, NY, USA.
khudabukhsh@mail.rit.edu

Mr. Sayantan Mahinder, Senior Research Scientist, Apple Inc., WA, USA.
smahinder@apple.com