Research Statement

Social media in today's world is a simmering cauldron of political messaging, personal branding by leaders and citizen engagement on the one hand, and misinformation and extreme speech driven by misogyny, jingoistic nationalism and religious/casteist bigotry on the other. While the former has piqued my curiosity in studying politics, I am alarmed by the effects of the latter on society. What drives me to pursue graduate studies at MIT is the potential for social media analysis to have a positive societal impact in emerging countries. Since 2014, India has seen a marked increase in instances of riots and mob violence, exacerbated by fake news and disinformation stories over digital media platforms like Facebook and Whatsapp. Through my recent work at Microsoft Research (MSR) on conspiracy theories and astroturfing on social media, and at IIT Delhi on the ideological biases prevalent in mass media, I have seen firsthand the power of our media ecosystem to shape the public narrative. In my key research projects described below, I have developed many of the Natural Language Processing, Data Visualisation and Machine Learning skills needed for impactful analysis of social networks.

My recent research projects have been centred around the use of social media platforms in mainstream politics, specifically around political brand building in India and in the US. I have been working with Prof. Joyojeet Pal as part of the Social Media Analysis Team at Microsoft Research (MSR), India. Firstly, I conducted a mixed-methods quantitative study to highlight themes of conspiracy and astroturfing across 33,000 politicians, 300 media houses and 3,000 journalists in the death of a popular film star (*Sushant Singh Rajput*) in India, which created massive uproar on the news cycle and social media. Using data from Twitter, YouTube, and an archive of debunked misinformation stories, we examined the drivers and consequences of social media outrage in this case, and contextualized the findings in the Indian socio-political backdrop. The retweet rates of media houses on Twitter indicated that the various commentators benefited from talking about the case as they received higher engagement than other topics during that time period. We also found that political entities had shaped the online discourse of the event by moving the early narrative that centered on depression and mental health, towards one of murder, mystery and conspiracy. Our early-stage findings[1] received widespread coverage in the press[2] and social media. The full paper is currently under review at CSCW 2021, where I am the primary author. This study bears witness to the complex web of interests and institutional actors involved in social media campaigns around the globe that can steer public discourse in a desired direction. It also shows us the need to understand the motivations of media houses, the ecosystem they inhabit and the role of astroturfing within misinformation.

Secondly, to understand the political dialogue on Twitter in terms of the topical choices of tweets, we developed a novel training method which learns the embedding of a politician by considering one tweet at a time, instead of concatenation. We used these embeddings to show that the local geography and party affiliation play an influential role in deciding what politicians write on social media in India and in the US. Our work is currently under review at ECIR 2021. Taking this idea further, I modelled the engagement (mentions/retweets) activity of politicians on Twitter as a big social graph with weighted links. The learnt embeddings show a strong correlation between the retweeting activity and the party affiliation of the politician. Further, I am currently analyzing the language divide in India on social media along with the content-retweet embeddings to understand the reasons why some states and parties show higher internal engagement than others. I am also simultaneously working on a written piece, examining the taxonomy of Indian politics on Twitter on a state-by-state basis. For each state, we discuss key politicians, their relationship with various influencers and journalists, engagement with the center and the way social media campaigns are organized for all major parties within the state, in our paper.

Working simultaneously on different projects has allowed me to build my research and people skills immensely. I got an opportunity to interact and brainstorm ideas with people from diverse academic backgrounds and research experiences, which helped me in inculcating efficient research practices. I was given complete ownership of the project and total flexibility in defining the research problem the way I wanted. The internship has allowed me to improve my academic writing skills, and frequent presentations helped me formulate research problems, and delineate research results clearly and concisely. I have also improved upon my result presentation by employing the use of creative visualization tools and libraries like `Pandas`, `Seaborn`, `Matplotlib` and `Plotly`, and `Charticulator`, amongst others.

Previously at IIT Delhi, I have worked with Prof. Aaditeshwar Seth for my Master's thesis to study the ideological biases prevalent in the Indian mass media on crucial economic and technology policies that affect our day to day lives. We built an end-to-end system that starts with a news article and parses it to obtain statements made by people in the article; on these statements, we applied a Recursive Neural Network based model to detect whether the statements express an ideological bias or not. Our results showed that the Indian mass media is ideologically biased, typically covering pro-policy statements much more than anti-policy statements and favoring a technology deterministic viewpoint more than the other side of the discourse. Our method can serve as a basis to contrast social media self-expression by prominent people with how the mass media portrays them. I took the lead in building up the research question and in paper-writing, and the paper[3] was accepted at the International Conference on Advances in Social Network Analysis and Mining (**ASONAM 2020**). This project showed how ideological biases in a media platform become an important apparatus in shaping public opinions, which piqued

[1] *Anatomy of a Rumour: Social media and the suicide of Sushant Singh Rajput* - S. Akbar, **A. Sharma**, H. Negi, A. Panda, J. Pal

[2] Press: Scroll, The Wire, India Today, The Quint, Mid-Day and NDTV; Social Media: Twitter, Twitter, Facebook, Instagram

[3] *Ideology Detection in the Indian Mass Media* - **A. Sharma**, N. Kaur, A. Sen, A. Seth - (ASONAM 2020)

my interest in media studies. I believe that our framework and findings can serve towards pushing the mass media towards greater self-regulation, enabling diversity in content publication, and educating the public about different viewpoints on key policies.

I had my first research experience in computational social science with Prof. Aaditeshwar Seth & Prof. Parag Singla at IIT Delhi, when I was working to analyze the irregularities in the onion and potato trading in India from 2006 to 2017. We curated a dataset of time series of wholesale prices and arrival volumes of the agricultural commodities at several village-level marketplaces, and retail prices of the commodities at the city centers. Our main task was to understand the reasons for fluctuations in food prices which can cause distress among both consumers and producers, and are often exacerbated by trading networks especially in developing economies. We came up with multivariate time series models using SARIMAX and LSTMs to forecast these prices accurately, and an anomaly detection and classification system to identify incidents of hoarding of stock by the traders. To improve the system further, I devised an innovative method to leverage the time-delays between different pairs of markets to account for the possible correlations between them appropriately. For validation, we got supporting facts from newspaper articles published during the time of the price hike. Our results point towards the possibility of building pricing models for agricultural commodities which can be used to reduce information asymmetries and to detect anomalies that can help regulate agricultural markets to operate more fairly. Not only did this project help me get exposure in dealing with large unstructured dataset of newspaper articles, but it also taught me different aspects of research, including ideation of the problem, literature review and the perseverance to explore new paths with active collaboration. I also got my first technical-writing experience in this project and we published our work[4] in **ACM COMPASS 2019**.

For exploring research in deep learning, I went for a semester-long research visit at **CSAIL (MIT)**, where I worked on through-wall Identity Recognition under the supervision of Prof. Dina Katabi. I trained a cross-modal teacher-student network for pose-estimation that transfers the visual knowledge of human pose using synchronized images and radio signals as a bridge. The student network generates confidence maps for all key-points of all people in the scene. Many motion-relevant GAIT features were extracted for each skeleton to uniquely model the identity of the individual. I used aggregation strategies on these extracted features for learning unsupervised representations robust to occlusions and environments. It was tested in through-wall identification scenarios for personalized and real-time sensing in smart homes as well. The stimulating environment in the lab and my interactions with fellow graduate students helped me realize the opportunities within the pursuit of further education. The opportunity to lead weekly meetings was also a learning experience, enabling me to develop soft skills to articulate my points well.

My introduction to the potential of AI was in a summer internship at the **University of Illinois at Urbana-Champaign (UIUC)** in the summer of 2017. I was supervised by Prof. Marianne Winslett on designing a novel attack framework to successfully reverse engineer the blueprints of an object from the factory floor using magnetometer side-channel data. Exploiting the design shortcomings of a 3D printer, I was able to experimentally show that the placement of motors on such a device can render it insecure. Integrating the entire reconstruction pipeline was a significant challenge since classical methods of segmentation on magnetometer spectrograms failed to generalize to new settings and environments. To build more adaptable systems, I started exploring the field of Machine Learning, which became a pivotal point in my career.

Besides research, I have also pursued software engineering internships at **Microsoft**, where my work included performance analysis of Azure storage systems (2018) and creating a contextual add-ins recommender for Outlook (2019). These internships allowed me to grow as a team player while working in a cross-team collaborative environment, and as an individual, when I learned to handle large codebases for applications that follow specific design considerations. After these development-oriented internships, I felt that my true allegiance lay with the opportunities to delve into the broader spectrum within academia. In addition to these technical skills, I have a strong academic record too. I secured an **All India Rank of 54** in the IIT Joint Entrance Examination (JEE) Mains 2015 from among a million candidates. My drive to excel continued in IIT Delhi, wherein during my Master's, I was ranked among the top 5 students in my degree program with a high grade point average of 9.24 on a 10 point scale.

My entire journey of projects and accomplishments led me to discover my research interests. Having said that, I envision my career as a researcher attempting to design technologies situated in the context of people, to solve challenging problems that communities face.

[4] *Price Forecasting & Anomaly Detection for Agricultural Commodities in India* - L. Madaan, **A. Sharma**, P. Khandelwal, S. Goel, P. Singla, A. Seth - Proceedings of the 2nd ACM SIGCAS Conference on Computing and Sustainable Societies, COMPASS '19