My research interests broadly lie in **Social Media Mining and Analysis**, integrating **Natural Language Processing (NLP)** and **Computational Social Science (CSS)**. The landscape of social media is highly dynamic, with users generating and consuming a diverse range of content. Various interest groups, including politicians, advertisers, and stakeholders, utilize these platforms to target potential users to advance their interests by adapting their messaging. This process, known as *microtargeting* [1, 2], relies on data-driven techniques that exploit the rich information collected by social networks about their users. Microtargeting is a double-edged sword; while it enhances the relevance and efficiency of targeted content, it also poses challenges, such as the risk of influencing user behavior and perceptions [3, 4], fostering echo chambers [5, 6, 7] and polarization [8, 9, 10]. Understanding these impacts is crucial for promoting healthy public discourse in the digital age and maintaining a cohesive society. **My work focuses on developing an organizing framework for a better understanding of microtargeting and activity patterns in social media on contentious topics**.

A significant challenge lies in understanding the messaging and how it changes depending on the targeted user groups. Another challenge arises when we do not know who the users are and what their motivations are for engaging with content. I address these challenges by developing computational approaches for (1) **characterizing user types and their motivations for engaging with content** [11, 12, 13, 14], (2) **analyzing the messaging based on topics relevant to the users and their responses to it** [15, 16, 17], and (3) **delving into the deeper understanding of the themes and arguments involved in the content** [18, 19, 20]. Currently, I am working on discovering latent themes and arguments of messaging leveraging pre-trained Large Language Models (LLMs). Based on this characterization, I aim to analyze demographic targeting and user engagement on social media platforms.

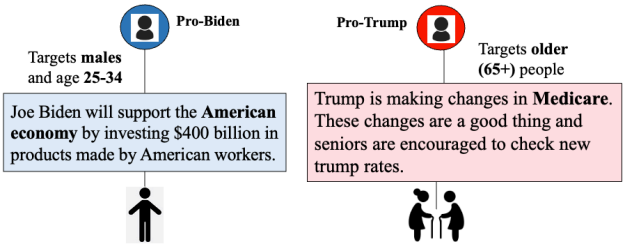

Figure 1: Example of Microtargeting

To analyze the impacts of microtargeting, understanding messaging from both the sender's and recipient's perspectives is essential. For the sender, we need to know what are their motivations. For the recipient, we need to know something about their demographic properties and interests, according to which we hypothesize that messaging would change. In the microtargeting example (Figure 1), Pro-Trump ads targeting the *older* (65+) population on *healthcare* issue adapt their messaging to emphasize the benefits of Trump's Medicare changes. Conversely, Pro-Biden ads targeting *males* and age group $25-34$ focus on *economy* issue, highlighting Biden's commitment to investing in American-made products.

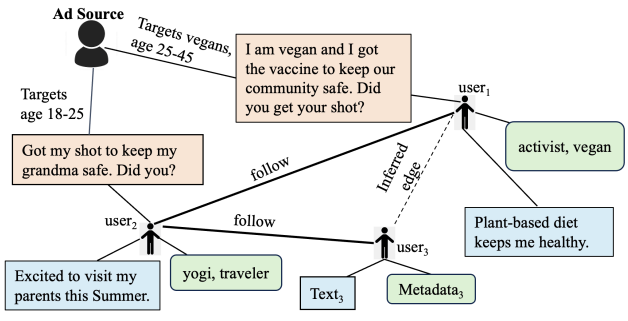

Figure 2: Information graph showing relations among users, advertisers (ad sources), texts (e.g., user-generated texts, ad texts), metadata (e.g., profile descriptions, demographic information).

In the context of social networks, I represent these interactions within an information graph. Figure 2 exemplifies my research's scope, depicting the complex interplay between users, advertisers, and content. It shows how users are interconnected and targeted by advertisers, a phenomenon I explore through data-driven analysis. We observe that $user_1$ and $user_2$ have *follow* connection, and both are *targeted* by the same ad source, which aims to get users vaccinated by adapting the messaging based on users' demographics and interests. My ap-

proaches predict the *user type* of $user_1$ as a vegan (`practitioner`), capture an *inferred edge* between $user_1$ and $user_3$ based on the social interactions and textual content, identify *stance* expressed in the ad text as (`pro-vaccine`) and the *argument* for it (`vaccine has not been tested on animal`). My work characterizes the *moral attitudes* expressed in the messaging (`loyalty`), and how different *entities* mentioned in it are perceived (`'I' am loyal to 'our community'`). My analysis also identifies the *issue* that the advertiser is interested in (`vaccination`). **My research is driven by characterizing users and messaging on social media**.

My research has made important contributions to computer science (CS), as recognized by my publications in prominent CS venues (ICWSM, NAACL, ACL, IEEEBigData, AIES), and awards (Purdue Graduate School Summer Research Grant).

Summary of Contribution

My goal is to comprehend how specific texts are directed to specific demographic attributes from specific interest groups. Working towards this goal, my research has significantly contributed in the following directions:

Characterizing user types and their motivations. The initial phase of my research focuses on comprehensively understanding users and their underlying motivations, whether practitioner-based or promotional. Gaining this understanding is crucial in reshaping our perspective on the content disseminated by these users.

As social media data is multifaceted, encompassing textual content, user interactions, and social connections, one challenge is integrating these different data views to construct a comprehensive representation of users. I introduce a joint embedding model for understanding social media users by incorporating their social and textual information [12]. This model is adept at capturing user types and motivations, specifically in the domains of **lifestyle choices, e.g., Yoga and Keto diet**.

The subsequent study [11] builds on this foundation by addressing the challenge of label scarcity. I propose a weakly supervised graph embedding-based framework that leverages the concept of social homophily [21] to infer user types from limited and weakly labeled data. This approach uses Expectation–maximization (EM) style iterative learning process to distinguish between nuanced user types, such as practitioners vs promoters, by analyzing their content shared on social media.

Characterizing messaging consumed by the users and their responses. In the era of pervasive social media, these platforms not only shape public opinion but also serve as an essential strategic tool for various stakeholders and interest groups, enabling them to achieve their goals by adapting messaging that resonates with specific demographic groups. Considering the profound impact of social media on public discourse, my research explores the realm of public campaigns and advocacy.

First, my work explores the decentralization of political messaging during the critical period of the **2020 U.S. presidential elections** [15]. Employing a weakly supervised learning model, my method identifies the stance and issue of political advertising, revealing how narratives are engineered to sway electoral outcomes by examining the relationship between users' responses (ad impressions) and election polls.

Further extending the scope of my analysis, I delve into the targeted messaging of **COVID-19 vaccination campaigns** on social media, uncovering the strategic adaptation of narratives to diverse demographic and geographic segments [17]. This study provides a minimally supervised multi-task learning framework with three different learning strategies to identify the theme and moral foundation of vaccine-related ads.

Additionally, my research investigates the landscape of **climate advertising** on social media [16]. At first, my work identifies the themes of the climate campaigns using an unsupervised approach. I propose a minimally supervised model soup [22] approach to identify stances combining themes of the content. By deploying the model, I analyze the stances within climate-related advertisements, offering unique insights into the interplay between advocacy, industry, and environmental activism.

Deeper understanding of the themes and arguments of content. Up until now, my approach to understanding messaging has relied on broader categories, i.e., a predefined set of labels, themes, and arguments that were fixed and established based on existing topics or theoretical frameworks, such as Moral Foundations Theory (MFT) [23]. However, this method often falls short of capturing the nuances of messaging. So, I shift my research focus to gaining a deeper understanding of messaging.

I introduce a holistic framework for analyzing social media posts about the **COVID-19 vaccine** [18]. It connects stance, reason analysis, moral foundation, and moral sentiment analysis at a granular entity level, successfully making reliable predictions with minimal supervision. This work studies how to model the dependencies between the different levels of analysis and incorporate **human insights** into the learning process.

Building on this, an interactive framework is developed that allows for the extraction of nuanced arguments from social media discussions, using the **COVID-19 vaccination debate** as a case study [19]. The framework balances automated methods with manual coding, streamlining the process while yielding more accurate and interpretable results.

My work further expands on the theme discovery process by proposing an interactive concept learning framework [20]. This framework redefines the nature of themes in text analysis to include generalized concepts validated by domain experts, moving beyond the limitations of traditional word co-occurrence-based approaches. This approach is validated through case studies on the **COVID-19 vaccine** and **immigration debates**.

Future Research Directions

I am passionate about solving meaningful **real-world problems**. Directions I am excited about include:

Opinion Mining. My vision for future research emphasizes understanding people's opinions and reasons for their decision choices. For instance, when policymakers want to introduce a new policy, it is crucial to predict the societal response: what would be people's opinion about it? Will the policy be welcomed or opposed? What motivates these attitudes? What would be the impact of that policy? My goal is to meticulously analyze these dynamics, providing valuable insights that can guide effective decision-making and foster positive societal change.

Integrating LLMs with a streamlined algorithmic process. My future research aims to understand the complex dynamics of social media and its impact on public discourse. In my initial work on content characterization, I focused on the level of topics that effectively captured patterns. However, I observed that the same topic could be the center of focus for distinctly different reasons, indicating a need for a more fine-grained view. Topics alone may not offer the level of granularity required for comprehensive analysis. Traditionally, this issue is addressed by having humans develop arguments and annotate data. Moving beyond this, my goal is to implement a **machine-in-the-loop** approach that aids in identifying relevant themes and arguments. I leverage the capabilities of powerful Large Language Models (LLMs) to characterize new themes and arguments, breaking down topics into finer-grained units. Furthermore, this approach includes identifying the mapping

between text and the discovered themes & arguments. This work goes beyond traditional human-centric data annotation and provides a faster generation of domain-specific labels.

Analyze demographic targeting and user engagement. Understanding demographic targeting and user engagement is vital due to its widespread implications across business, technology, policy-making, public health, society, and culture. It not only aids in the optimization of digital marketing strategies but also contributes to a deeper understanding of user behavior and societal trends. My goal is to analyze demographic targeting and user engagement on social media messaging based on the newly discovered themes and arguments. Furthermore, I am interested in highlighting ethical considerations in demographic targeting, particularly regarding the potential for bias or discrimination in targeted content.

Future Collaboration. I intend to conduct technical research in collaboration with researchers from diverse backgrounds. During my M.Sc., I gained research experience in Bioinformatics and Computational Biology. I developed computational methods for (1) improved region-specific assembly for the human genome [24, 25, 26] and (2) quantifying the twist of β-strands in the secondary structure of proteins [27, 28].
My work is inherently interdisciplinary. Given my previous background in **Bioinformatics** and **Computational Biology** and the current focus on **NLP** and **CSS**, there are several opportunities for meaningful collaboration with other fields:

- **Health Informatics:** Applying NLP to extract and analyze information from clinical notes, medical journals, and research papers. This collaboration could lead to advancements in drug discovery, understanding disease patterns, and improving patient care.
- **Psycholinguistics:** Exploring how language reflects cognitive processes and mental health status. This might involve analyzing social media texts to detect patterns indicative of mental health issues, aiding in early diagnosis and treatment strategies.
- **Environmental Science:** Analyzing social media and other textual data to gauge public opinions and attitudes towards environmental issues, climate change, and sustainability practices.
- **Sociolinguistics:** Studying language use on social media to understand societal trends, cultural differences, and social dynamics. Such collaboration can offer insights into public opinion, social movements, and cultural shifts.
- **BioNLP:** Leveraging NLP techniques to decipher complex patterns in genetic and protein sequences. The ability to decode and interpret these biological sequences through the lens of computational linguistics opens new avenues for understanding molecular biology.
- **Public Health Surveillance**: Integrating CSS methodologies to analyze social media and online content for tracking public health trends, disease outbreaks, and public perceptions about health-related issues. This approach can provide early warning signals for public health emergencies or insights for health campaigns.

References

[1] Eitan D Hersh. *Hacking the electorate: How campaigns perceive voters*. Cambridge University Press, 2015.

[2] Oana Barbu. Advertising, microtargeting and social media. *Procedia-Social and Behavioral Sciences*, 163:44–49, 2014.

[3] Filipe N Ribeiro, Koustuv Saha, Mahmoudreza Babaei, Lucas Henrique, Johnnatan Messias, Fabricio Benevenuto, Oana Goga, Krishna P Gummadi, and Elissa M Redmiles. On microtargeting socially divisive ads: A case study of russia-linked ad campaigns on facebook. In *ACM FAccT*, 2019.

[4] Sanne Kruikemeier, Minem Sezgin, and Sophie C Boerman. Political microtargeting: relationship between personalized advertising on facebook and voters' responses. *Cyberpsychology, Behavior, and Social Networking*, 19(6):367–372, 2016.

[5] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 World Wide Web Conference*, pages 913–922. International World Wide Web Conferences Steering Committee, 2018.

[6] Walter Quattrociocchi, Antonio Scala, and Cass R Sunstein. Echo chambers on facebook. *Available at SSRN 2795110*, 2016.

[7] Kathleen Hall Jamieson and Joseph N Cappella. *Echo chamber: Rush Limbaugh and the conservative media establishment*. Oxford University Press, 2008.

[8] Julie Jiang, Emily Chen, Shen Yan, Kristina Lerman, and Emilio Ferrara. Political polarization drives online conversations about covid-19 in the united states. *Human Behavior and Emerging Technologies*, 2(3):200–211, 2020.

[9] Frederik Zuiderveen Borgesius, Judith Möller, Sanne Kruikemeier, Ronan Ó Fathaigh, Kristina Irion, Tom Dobber, Balazs Bodo, and Claes H de Vreese. Online political microtargeting: Promises and threats for democracy. *Utrecht Law Review*, 14(1):82–96, 2018.

[10] Ann Bostrom, Gisela Böhm, and Robert E O'Connor. Targeting and tailoring climate change communications. *Wiley Interdisciplinary Reviews: Climate Change*, 4(5):447–455, 2013.

[11] **Tunazzina Islam** and Dan Goldwasser. Twitter user representation using weakly supervised graph embedding. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 358–369, 2022.

[12] **Tunazzina Islam** and Dan Goldwasser. Analysis of twitter users' lifestyle choices using joint embedding model. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 15, pages 242–253, 2021.

[13] **Tunazzina Islam** and Dan Goldwasser. Do you do yoga? understanding twitter users' types and motivations using social and textual information. In *2021 IEEE 15th International Conference on Semantic Computing (ICSC)*, pages 362–365. IEEE, 2021.

[14] **Tunazzina Islam** and Dan Goldwasser. Does yoga make you happy? analyzing twitter user happiness using textual and temporal information. In *2020 IEEE International Conference on Big Data (Big Data)*, pages 4241–4249. IEEE, 2020.

[15] **Tunazzina Islam**, Shamik Roy, and Dan Goldwasser. Weakly supervised learning for analyzing political campaigns on facebook. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 17, pages 411–422, 2023.

[16] **Tunazzina Islam**, Ruqi Zhang, and Dan Goldwasser. Analysis of climate campaigns on social media using bayesian model averaging. In *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '23, page 15–25, New York, NY, USA, 2023. Association for Computing Machinery.

[17] **Tunazzina Islam** and Dan Goldwasser. Understanding covid-19 vaccine campaign on facebook using minimal supervision. In *2022 IEEE International Conference on Big Data (Big Data)*, pages 585–595. IEEE, 2022.

[18] María Leonor Pacheco*, **Tunazzina Islam***, Monal Mahajan, Andrey Shor, Ming Yin, Lyle Ungar, and Dan Goldwasser. A holistic framework for analyzing the covid-19 vaccine debate. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 5821–5839, 2022.

[19] Maria Leonor Pacheco, **Tunazzina Islam**, Lyle Ungar, Ming Yin, and Dan Goldwasser. Interactively uncovering latent arguments in social media platforms: A case study on the covid-19 vaccine debate. In *Proceedings of the Fourth Workshop on Data Science with Human-in-the-Loop (Language Advances)*, pages 94–111, 2022.

[20] Maria Leonor Pacheco, **Tunazzina Islam**, Lyle Ungar, Ming Yin, and Dan Goldwasser. Interactive concept learning for uncovering latent themes in large text collections. In Anna Rogers, Jordan Boyd-Graber, and Naoaki Okazaki, editors, *Findings of the Association for Computational Linguistics: ACL 2023*, pages 5059–5080, Toronto, Canada, July 2023. Association for Computational Linguistics.

[21] Miller McPherson, Lynn Smith-Lovin, and James M Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444, 2001.

[22] Mitchell Wortsman, Gabriel Ilharco, Samir Ya Gadre, Rebecca Roelofs, Raphael Gontijo-Lopes, Ari S Morcos, Hongseok Namkoong, Ali Farhadi, Yair Carmon, Simon Kornblith, et al. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. In *International Conference on Machine Learning*, pages 23965–23998. PMLR, 2022.

[23] Jonathan Haidt and Jesse Graham. When morality opposes justice: Conservatives have moral intuitions that liberals may not recognize. *SJR*, 2007.

[24] **Tunazzina Islam**, Desh Ranjan, Eleanor Young, Ming Xiao, Mohammad Zubair, and Harold Riethman. Rextal: Regional extension of assemblies using linked-reads. In *Bioinformatics Research and Applications: 14th International Symposium, ISBRA 2018, Beijing, China, June 8-11, 2018, Proceedings 14*, pages 63–78. Springer, 2018.

[25] **Tunazzina Islam**, Desh Ranjan, Mohammad Zubair, Eleanor Young, Ming Xiao, and Harold Riethman. Analysis of subtelomeric rextal assemblies using quast. *IEEE/ACM transactions on computational biology and bioinformatics*, 18(1):365–372, 2019.

[26] Eleni Adam, **Tunazzina Islam**, Desh Ranjan, and Harold Riethman. Nanopore guided assembly of segmental duplications near telomeres. In *2019 IEEE 19th international conference on bioinformatics and bioengineering (BIBE)*, pages 60–65. IEEE, 2019.

[27] **Tunazzina Islam**, Michael Poteat, and Jing He. Quantification of twist from the central lines of β-strands. *Journal of Computational Biology*, 25(1):114–120, 2018.

[28] **Tunazzina Islam**, Michael Poteat, and Jing He. Analysis of β-strand twist from the 3-dimensional image of a protein. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics*, pages 650–654, 2017.