

# Identity Decolonization amid the Coloniality of Computing

by

**Dipto Das**

M.S., University of Colorado Boulder, 2023

M.S., Missouri State University, 2019

B.S.Engg., Bangladesh University of Engineering and Technology, 2017

A thesis submitted to the  
Faculty of the Graduate School of the  
University of Colorado in partial fulfillment  
of the requirements for the degree of  
Doctor of Philosophy  
Department of Information Science

2024

Committee Members:

Dr. Bryan Semaan, Chair

Dr. Casey Fiesler

Dr. Jed Brubaker

Dr. Lilly Irani

Dr. Syed Ishtiaque Ahmed

Das, Dipto (Ph.D., Information Science)

Identity Decolonization amid the Coloniality of Computing

Thesis directed by Prof. Dr. Bryan Semaan

Through colonialism, external forces can alter and shift social structures and practices. It causes trans-generational, often normalized, invisible, and profound marginalization of the collective identities of local and indigenous populations. Decolonization is the resisting and undoing of colonial impacts. It's the process of reforming a society's social, cultural, economic, and political structure to reflect more of the local and indigenous values than that of the colonial rulers. While sociotechnical systems can support the identity work of marginalized communities, social computing researchers have also discussed how these systems can cause harassment, exclusion, and other kinds of harm, impose values, and exhibit colonial impulses. While previous works in HCI and social computing have focused on several marginalized communities, there is a dearth of literature on the identity works of colonially marginalized communities using ICT platforms. My research contributes to understanding how colonialism marginalized people in the Global South across various dimensions of identity (e.g., race, gender, sexuality, religion, caste, nationality), how sociotechnical systems reinstate colonial structures and values, and how computing platforms both support and impede colonially marginalized communities' identity expression and performances.

My dissertation examines Bengali ethnolinguistic communities' identity work, expression, and performances in ICT spaces, who have been marginalized by foreign powers for ages and are under-represented in computing. Building on decolonial and postcolonial perspectives with a historicist sensibility, in this dissertation, my mixed-method empirical studies on various platforms highlight users' agency, the role of content moderation, algorithms, datasets, online communities, and their policies seek to understand how the previously colonized Bengali people decolonize their identities through these platforms and how the embedded coloniality of different components of these systems impede their identity work and expression. In the first two studies using trace ethnography, I

sought to understand how people use the predominantly text-based Q&A site Bengali Quora to collaboratively reclaim narrative agency as a form of identity decolonization and how the governance approach of biased human moderators, collective surveillance, and algorithmic coloniality impedes that process. I built on these studies in terms of multimodality, ways of communication, and moderation approaches. Through semi-structured interviews with content creators on YouTube, where the interaction among users, contrary to the text-based collaborative website Quora, is multimodal and content creator-centric. This study explores YouTubers' motivation and strategies for making videos toward decolonial discourse and how they navigate different challenges in their work. For the final study, I explored the feasibility of automated content moderation by examining one of its common components, like NLP (e.g., sentiment analysis) tools, algorithms, and datasets. In doing so, my work altogether informs the broader social computing literature on identity, content moderation, fairness and bias, social justice, and ICT for development.

## Dedication

This dissertation is dedicated to my mother, Dipa Das, and my late father, Nirmal Das.

In 2013, the morning of the day my father passed away, I was wandering around in the house, having just completed the first semester at the university. While sick, my father told me “Go and study a little bit.” I told him that my semester just ended, so there was not much study to do. But he told me to at least sit down with a book. He often told me the stories he heard at the small shop where he worked in a bazaar about people going abroad for higher studies. He asked me how much it could have cost for me to do the same. It was like daydreaming in a family that lived from paycheck to paycheck. After my father passed away in the afternoon that day, my mother, my younger sister, and I were crying in grief. When people came to console us, I still remember my mother telling them, “My son won’t be able to continue his studies now.” She had just lost the love of her life, and one of the first things she was concerned about was my education! That’s how much she values my education. On the days it felt difficult, I strongly believe that her prayers and motivation were what kept me going. My mother’s priorities and my father’s wishes for my higher studies absolutely inspired me to pursue a master’s and then a doctoral degree in the US.

To whom else can I dedicate this but to the Merciful God’s greatest blessing to me, my loving parents, who taught me to love learning and sacrificed everything for that purpose?

## Acknowledgements

I am thankful to the Almighty God for granting me the opportunity to pursue doctoral study and the ability to complete the dissertation research.

I want to express my gratitude to my dissertation committee members, Drs. Jed Brubaker, Casey Fiesler, Lilly Irani, Syed Ishtiaque Ahmed, and Bryan Semaan for their assistance and mentorship in my dissertation research. I am especially grateful to my advisor, Dr. Bryan Semaan, who went above and beyond to constantly support and guide me in this journey. His kind gestures, such as including me while moving to CU Boulder, advising me whenever and whatever I needed, and exploring various ways to help me during my difficult times, demonstrated his care for my scholarly development and well-being. A special thanks goes to Dr. Ahmed for inspiring me to pursue research since I was an undergraduate student. I also deeply appreciate my collaborators, Drs. Shion Guha and Carsten Østerlund for their invaluable feedback and guidance in my research.

I am forever indebted to my mother, Dipa Das, and my sister, Dipty Das, for their constant support and motivation. I respectfully remember my late father, Nirmal Das, who passionately wanted me to pursue higher studies. I am grateful for the support I received from my other family members. I am thankful to have been blessed with wonderful teachers since my childhood schools who still guide me in various ways. I thank the friends and strangers I met in the United States who made this country feel like home to me. Their value to me increases with time.

I thank my research assistants for their support and the interview and survey participants for contributing their insights and time to my research. This research was partially supported by NSF CRII#1657429, CMCI research funds, and travel grants from ACM, NSF, and Meta.

## Contents

<b>Chapter</b>	
<b>1</b>	<b>Introduction</b> <span style="float: right;"><b>1</b></span>
1.1	Research Approach and Methods . . . . . 3
1.2	Research Overview and Structure of Dissertation . . . . . 4
1.2.1	Chapter 2: Literature Review . . . . . 5
1.2.2	Chapter 3: Research Site and Researcher Positionality . . . . . 5
1.2.3	Chapter 4: Methods . . . . . 5
1.2.4	Chapter 5: Collaborative Reclamation of Narrative Agency . . . . . 5
1.2.5	Chapter 6: Governance Shaping Platform Identity . . . . . 6
1.2.6	Chapter 7: Content Creation to Reimagine Transnational Communities . . . 6
1.2.7	Chapter 8: Postcolonial Politics of Content Creation . . . . . 7
1.2.8	Chapter 9: Coloniality of Algorithmic Tools and Datasets . . . . . 7
1.2.9	Chapter 10: Toward Decolonization amid the Coloniality of Computing . . . 8
1.2.10	Chapter 11: Conclusion . . . . . 8
<b>2</b>	<b>Literature Review</b> <span style="float: right;"><b>9</b></span>
2.1	The Relationship among Identity, Marginalization, and Colonialism . . . . . 9
2.1.1	How Colonialism Marginalized Collective Identities . . . . . 10

2.1.2	How Colonialism Impacted Social Identities in Bengali Communities in the Indian Subcontinent . . . . .	12
2.2	Decolonization as Regaining Control Over Identity . . . . .	14
2.3	Technology and Digital Identity . . . . .	16
<b>3</b>	<b>Research Site and Researcher Positionality</b>	<b>18</b>
3.1	Bengali People and Their History with Colonization . . . . .	19
3.2	Researcher Reflexivity . . . . .	23
<b>4</b>	<b>Methodological Overview</b>	<b>25</b>
4.1	Phase 1: Trace Ethnography of Quora . . . . .	25
4.2	Phase 2: Interviews with YouTubers . . . . .	27
4.3	Phase 3: Audit of Algorithmic Systems . . . . .	29
<b>5</b>	<b>Collaborative Reclamation of Narrative Agency</b>	<b>31</b>
5.1	Background . . . . .	32
5.1.1	Colonialism as Shaping Narrative Agency: Losing Control Over Identity . . .	32
5.1.2	Decolonization as Re-Establishing Narrative Agency . . . . .	35
5.1.3	Reclaiming Narrative Agency with Technology . . . . .	36
5.2	Methodological Details . . . . .	38
5.2.1	Data Collection . . . . .	38
5.2.2	Data Analysis . . . . .	45
5.2.3	Limitations . . . . .	46
5.3	Results . . . . .	47
5.3.1	Rediscovery and Recovery: Collaborative Conceptualization and Healing from Collective Colonial Trauma . . . . .	47
5.3.2	Mourning to Continue Healing from Collective Trauma: Speculative Identity Play . . . . .	54

5.3.3	Dreaming Towards Unity: Reconfiguring Geopolitical Relations . . . . .	56
5.3.4	Commitment to Commonality through Anti-Categorical Identity Framing . .	60
5.3.5	Action towards Repairing Sociocultural Structures: BnQuora as an Identity Mirror . . . . .	62
5.4	Discussion . . . . .	64
<b>6</b>	<b>Governance Shaping Platform Identity</b>	<b>68</b>
6.1	Background . . . . .	69
6.1.1	Performative Identity and Linguistic Marginalization . . . . .	69
6.1.2	Governance as Surveillance: Towards Platform Identity . . . . .	71
6.2	Methodological Details . . . . .	74
6.2.1	Data Collection . . . . .	74
6.2.2	Data Analysis . . . . .	76
6.3	Results . . . . .	77
6.3.1	Sociotechnical Mechanisms of Governance on BnQuora . . . . .	77
6.3.2	Panoptic Performativity and Platform Identity on BnQuora . . . . .	80
6.4	Discussion . . . . .	91
<b>7</b>	<b>Content Creation to Reimagine Transnational Communities</b>	<b>95</b>
7.1	Background . . . . .	96
7.1.1	Nationalism as a Collective Identity . . . . .	97
7.1.2	How Colonialism Has Shaped Nationalism . . . . .	98
7.1.3	Institutions in Shaping Hegemonic Nationalist Discourse . . . . .	100
7.1.4	Reimagining Communities through Decolonial Discourse on Online Platforms	102
7.2	Methodological Details . . . . .	105
7.2.1	Recruitment . . . . .	105
7.2.2	Interviews . . . . .	107
7.2.3	Data Analysis . . . . .	108

7.3	Results . . . . .	109
7.3.1	Nationalism’s Impact on Cultural Community: The “Inner” Domain of Cultural Identity . . . . .	110
7.3.2	Nationalism Defined through Institutions and Establishments: The “Outside” Domain of Institutionalization . . . . .	122
7.4	Discussion . . . . .	132
7.4.1	Modularization and Polarization in Online Discourse . . . . .	133
7.4.2	YouTube’s Video Modality for Inclusive Decolonial Discourse . . . . .	135
<b>8</b>	<b>Postcolonial Politics of Content Creation</b>	<b>137</b>
8.1	Background . . . . .	138
8.2	Methodological Details . . . . .	140
8.2.1	Recruitment and Interviews . . . . .	140
8.2.2	Data Analysis . . . . .	140
8.3	Results . . . . .	140
8.3.1	Postcolonial Tensions in Identities and Relationships . . . . .	141
8.3.2	Postcolonial Structures Mediated through Technology . . . . .	149
8.4	Discussion . . . . .	159
<b>9</b>	<b>Coloniality of Algorithmic Tools and Datasets</b>	<b>161</b>
9.1	Background . . . . .	163
9.1.1	Expressions of Social Identity through Language and Technology . . . . .	164
9.1.2	Algorithmic Bias Deconstruction in Computing Systems . . . . .	166
9.2	Methodological Details . . . . .	170
9.2.1	Developing Bengali Identity-based Bias Evaluation Dataset (BIBED) . . . . .	171
9.2.2	Identifying Bengali Sentiment Analysis Tools . . . . .	180
9.2.3	Identifying Bengali Sentiment Analysis Datasets . . . . .	182
9.2.4	Identifying Language Models for the Bengali Language . . . . .	182

9.2.5	Experiment Setup for Algorithmic Audit . . . . .	184
9.2.6	Environmental Impact . . . . .	189
9.3	Results . . . . .	190
9.3.1	BSA tools' Presumed Universality and Reductionist Representation . . . . .	190
9.3.2	Colonial Hierarchies and Politics of Design . . . . .	194
9.3.3	How Do Different Combinations of BSA Datasets and Language Models Lead to Biased Tools? . . . . .	200
9.4	Discussion . . . . .	203
9.4.1	Inconsistencies in Sentiment Analysis Tools' Outputs . . . . .	204
9.4.2	Codification of Implicit Expression of Identities in Sentiment Analysis . . . . .	205
9.4.3	Exploring Downstream Effects of Bias in Sentiment Analysis Through the Context of Tool Development and Content Moderation . . . . .	206
9.4.4	Collaboration among Developers of Diverse Demographic Backgrounds . . . . .	208
<b>10</b>	<b>Toward Decolonization amid the Coloniality of Computing</b>	<b>211</b>
10.1	Re-conceptualizing Colonization and Contextualizing Its Impacts . . . . .	212
10.1.1	Implications for Explorations of Disruption . . . . .	212
10.1.2	Double Consciousness of Religious Communities and the Veil in Transnational Communication . . . . .	214
10.2	Reflection on Methodologies and Implications . . . . .	216
10.2.1	Can All Marginalized People Speak?: Expectation for Sampling . . . . .	216
10.2.2	Sociomaterial Considerations in Trace Data Collection . . . . .	217
10.2.3	Quantitative Methods in Critical Scholarship . . . . .	222
10.2.4	Ethical Consideration in Data Curation . . . . .	223
10.3	Implications for HCI and CSCW Research . . . . .	225
10.3.1	Postcolonial Sociomateriality of Online Interaction . . . . .	225
10.3.2	Invitation to Critically Study Nationalism in CSCW . . . . .	227

10.3.3	Historicist Understanding for Cross-Cultural Computing . . . . .	228
10.3.4	Call for Engineering Activism in Critical HCI . . . . .	230
<b>11</b>	<b>Conclusion</b>	<b>232</b>
	<b>Bibliography</b>	<b>235</b>
	<b>Appendix</b>	
<b>A</b>	<b>Quoras API</b>	<b>278</b>
A.1	Implementation . . . . .	278
A.2	Usage . . . . .	279
<b>B</b>	<b>BIBED Preparation Details</b>	<b>281</b>
<b>C</b>	<b>Detailed Results of Algorithm Audits</b>	<b>284</b>
C.1	Numerical Results from Auditing BSA Tools . . . . .	284
C.2	Numerical Results of Auditing Different Combinations of BSA Datasets and Lan- guage Models . . . . .	297
<b>D</b>	<b>List of Developed Artifacts</b>	<b>301</b>

## List of Tables

### Table

5.1	List of keywords and phrases. The search keywords/phrases are sorted in the alphabetic order of the second column for easier lookup for non-Bengali speaking readers.	39
6.1	List of keywords and phrases. The search keywords/phrases are not sorted in any particular order. . . . .	75
7.1	Demographic information of the participants . . . . .	107
9.1	Identity dimensions and corresponding categories focused in this study. . . . .	170
9.2	Examples of sentence pairs from Gender, Religion, and Nationality-based EBE datasets. Translations are shown inside parentheses. . . . .	173
9.3	Sentences using common names and kinship terms in different religious communities.	175
9.4	An English sentence’s Bengali translations resembling Bangladeshi and Indian dialects.	176
9.5	Number of sentences included in the dataset from different stages of compilation. . .	178
9.6	Bengali sentiment analysis tools examined in this chapter (T1 is from PyPI and T2-T13 are from GitHub). In “Developer Demographics” column, we used icons to represent identity categories: female (♀), male (♂), Hindu (ॐ), Muslim (☪), Bangladeshi (🇧🇩), and Indian (🇮🇳). . . . .	181
9.7	Bengali sentiment analysis datasets examined in this chapter. . . . .	183

9.8	BSA tools' bias toward gender identity categories grouped by their developers' gender identities. . . . .	199
9.9	BSA tools' toward religion-based bias grouped by their developers' religious identities.	199
9.10	BSA tools' nationality-based bias grouped by their developers' national identities. .	199
9.11	Fine-tuned BSA models' bias toward gender identity categories grouped by the BSA datasets' developers' gender identities. <i>Dxm</i> and <i>DxB</i> respectively indicate the fine-tuned models resulting from training <i>multilingualBERT</i> and <i>BanglaBERT</i> using the BSA dataset <i>Dx</i> . . . . .	201
9.12	Fine-tuned BSA models' bias toward gender identity categories grouped by the BSA datasets' developers' religious identities. . . . .	201
9.13	Fine-tuned BSA models' bias toward gender identity categories grouped by the BSA datasets' developers' national identities. . . . .	201
A.1	<b>quoras</b> Data collection functions. . . . .	280
B.1	Examples of female and male names associated with being Bengali Hindu and Bengali Muslim. . . . .	281
B.2	Pairs of noun phrases representing kinship with a female or a male person in Bengali Hindu and Bengali Muslim communities. . . . .	282
B.3	Sentence templates used in generating name-based IBE dataset. . . . .	283
B.4	Different words with same meaning in Bangladeshi and Indian colloquial vocabulary.	283
C.1	Results of Kruskal-Wallis analysis of variance of different sentiment analysis tools' mean scores for (a) female, (b) male, (c) Hindu, (d) Muslim, (e) Bangladeshi, and (e) Indian identity categories. Here, the larger the H-statistic values are, the stronger the evidence against the null hypotheses. . . . .	284
C.2	Conover-Iman test across BSA tool pairs for Female identity category . . . . .	285
C.3	Conover-Iman test across BSA tool pairs for Male identity category . . . . .	286

C.4	Conover-Iman test across BSA tool pairs for Hindu identity category . . . . .	287
C.5	Conover-Iman test across BSA tool pairs for Muslim identity category . . . . .	288
C.6	Conover-Iman test across BSA tool pairs for Bangladeshi identity category . . . . .	289
C.7	Conover-Iman test across BSA tool pairs for Indian identity category . . . . .	290
C.8	Results of Wilcoxon signed-rank test between female and male gender-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{female} = \mu_{male}$ ) and alternative hypotheses in different directions . . . . .	291
C.9	Results of Wilcoxon signed-rank test between Hindu and Muslim religion-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{Hindu} = \mu_{Muslim}$ ) and alternative hypotheses ( $H_{RA}$ ) in different directions . . . . .	293
C.10	Results of Wilcoxon signed-rank test between Bangladeshi and Indian nationality-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{Bangladeshi} = \mu_{Indian}$ ) and alternative hypotheses ( $H_{NA}$ ) in different directions . . . . .	295
C.11	Power of $\chi^2$ and Wilcoxon/t-tests comparing sentiment labels and scores assigned for different identity categories by fine-tuned models using different combinations of datasets and language models. . . . .	297
C.12	Quantified Bias Metrics (average PCM and PCR) in ten data splits. . . . .	299

## List of Figures

### Figure

- 4.1 A screenshot of a public BnQuora Q&A thread (<https://qr.ae/pGPOSZ>) accessed from the first author’s account. The thread currently has 16 answers, only two of them are shown in the image as examples. Text on the right sides of the arrows describe the component (in bold) and translated short text (in italics). . . . . 26
- 6.1 Panoptic Platform depiction of BnQuora. . . . . 93
- 9.1 An example entry from our dataset. . . . . 179
- 9.2 (a) Fine-tuning multilingualBERT or BanglaBERT using different BSA datasets (e.g., *Dx*). (b) Auditing fine-tuned *Dx*-multilingualBERT or *Dx*-BanglaBERT models’ gender, religion, and nationality-based biases. . . . . 186
- 9.3 Experimental setup for comparing different BSA tools’ outputs for fixed identity category (e.g., female) . . . . . 191
- 9.4 Comparing sentiment scores for an identity (e.g., Bangladeshi) expressed explicitly and implicitly (visualized using solid and dashed lines, respectively). . . . . 193

9.5	Experimental setup for comparing sentiment scores for different categories under an identity dimension. From top-middle-bottom: the schematics represent setups for gender (female-male), religion (Hindu-Muslim), and nationality (Bangladeshi-Indian), and the similarity of sentence pairs is indicated by the icon <i>lorem</i> . We consistently ordered the categories in each pair alphabetically. . . . .	195
9.6	Distributions of scores from different BSA tools for sentences expressing different identity categories. . . . .	196
9.7	Heatmap showing the directions of biases of the fine-tuned models based on Positive Classification Rate, i.e., in how many iterations a particular combination of <code>multilingualBERT/BanglaBERT</code> with different BSA datasets more frequently classified an identity category as positive. . . . .	202
10.1	Trace data collection process in studies in BnQuora. . . . .	220
A.1	Example <b>quoras</b> workflow in an iPython notebook. . . . .	280

## Chapter 1

### Introduction

Identity has been central to social computing scholarship and public initiatives toward diversity and inclusion. People’s identity—how they see themselves and want others to see them as physical and social beings [221], is both an individuated and social construct. Identity is often shaped by their membership in different groups [530]. Identity expression as parts of various social groups is not fixed but co-constructed through everyday performances, values, and practices of these communities [100]. While people want to express and enact their identities freely and without harm, they are often marginalized or pushed to the periphery of society based on their identity across various dimensions such as race, gender, language, sexual orientation, nationality, religion, caste, educational and socioeconomic status [192]. Human-computer interaction (HCI) and computer-supported cooperative work (CSCW) researchers have studied technology’s role in supporting and impeding the identity expression of marginalized communities, based across identity intersections such as race, gender, and sexual orientation, predominantly in Western and Global North contexts [252, 320, 610]. However, a dearth of scholarship focuses on identity in the Global South. As identity mediates from people’s everyday social interaction to geopolitical relationships, studies on identities in the Global South would help the fields of HCI and social computing understand people’s experiences and perspectives in the majority world. What is also often missing in these studies is the genealogy of how certain identities have been historically normalized and prioritized at the cost of marginalizing others. In recent years, scholars have highlighted the importance

of adopting a historicist sensibility to inform CSCW research [534]. Historicism can offer a deeper understanding of how perceptions of different identities have developed in our societies and how different communities are marginalized based on their identity.

Colonialism refers to the policies and practices where external powers migrate to other lands and alter the social, cultural, political, and economic structures and, thus, identities of local and indigenous populations [352]. Given its impact on myriad dimensions of human identity like race [199], gender [356], sexuality [362], religion [101], caste [421], nationality [111], and diaspora [161], I<sup>1</sup> argue that studying the marginalization of identity in relation to colonialism would help understand the sociohistoric entanglement of marginalized identities, especially in the Global South, with various emerging forms of technology. From human-classification schemata to modern computing spaces, scholars have highlighted myriad “colonial impulses” (explained in Chapter 2) of computing technologies [17, 180] and how, by virtue of being designed in Western contexts, as sociotechnical systems migrate and travel to other, especially non-Western contexts, they reanimate coloniality by inflicting Western values and ways of being on others [282]. Therefore, postcolonial scholars who study the impacts of colonization and decolonial scholars who explore ways to resist Western supremacy over its colonized subjects have conceptualized colonization as the long-term, normalized, and often invisible mechanism of cultural imposition and marginalization of people’s identities based on race, ethnicity, nationality, religion, and language. Considering the sociohistoric and technological dimensions of colonial influence on identity, my research agenda and this dissertation are guided by the following two broad questions.

- **RQ1:** How can technology *support* colonially marginalized communities’ identity expression and performance?

---

<sup>1</sup> While I have led the studies presented in this dissertation, I recognize the collaborative nature of this research. Therefore, except for my positionality statement in Chapter 3, I have used the royal “we” pronoun throughout this document, such as in the methods and results chapters (4-9). When I say “we”, I am acknowledging the contributions of my advisor, co-authors, mentors, and research team who have provided invaluable support and feedback. While I collected data, conducted interviews, surveys, and experiments, and analyzed data to reach the interpretations and my own arguments presented in this dissertation, I sincerely appreciated their help in spreading my call for recruiting participants, recommendations about relevant literature, and support to overcome miscellaneous challenges.

- **RQ2:** How can technology *impede* colonially marginalized communities' identity expression and performance?

To address these questions, my research centers on the online practices of the Bengali communities who are native to the Indian subcontinent. Using historical materials, I seek to understand how colonial power marginalized Bengali culture and shaped Bengali identity. In this dissertation, I explore whether and how online community spaces (e.g., video-sharing and Q&A sites) provide opportunities for colonially marginalized Bengali communities to engage in identity decolonization—negotiate and think about their identities to resist colonial influences on their local and indigenous selves. At the same time, as computing systems reify and reinforce colonial values, control, and structures, I critically study whether and how the Bengali communities face exclusion and further marginalization in these spaces by different sociotechnical components of these platforms, such as algorithms, datasets, governance, and content moderation. Overall, I study the entanglement between technology's capacity to empower marginalized communities in expressing their identities and the inherent colonial biases ingrained within computing, framing my work as an exploration of Bengali users' journey toward identity decolonization amid the coloniality of computing.

## 1.1 Research Approach and Methods

My research area is at the intersection of HCI, CSCW, social computing, and ICT for development (ICTD). As will be unpacked in the literature review (chapter 2), I draw from the literature on decolonial and postcolonial computing, identity studies, and online communities to guide my work. Over the years, a significant body of CSCW and HCI research has studied how people perform their identity in computing spaces, what constitutes digital identity, and how technology supports and impedes the expression of marginalized identities. However, many questions remain unanswered about identity in computing spaces. Relatively little scholarship in HCI and CSCW has explored Indigenous and colonially marginalized identities [41, 152, 586], especially in the context of the Global South. The Indian subcontinent is a site of prolonged colonization

with diverse identities—uncovering the impacts of colonization, indigeneity, nationhood based on language, religion, geopolitical boundaries, and more poses increased complexity that demands a critical and nuanced understanding of social, historical, cultural, and political factors. Moreover, as a decolonial effort, I actively turn to the scholarship from the Global South in HCI and beyond to critique, interrogate, and augment existing computing literature, which often presents West-centric perspectives [17].

To address my research questions, I employ a mixed-method approach that includes qualitative methods like semi-structured interviews and trace ethnography and quantitative methods like surveys, natural language processing (NLP), and experimental design. Through my investigation of multiple platforms and systems (e.g., Quora, YouTube, and NLP tools), I seek to understand the Bengali users' experience with sociotechnical systems of various scaffolds and affordances. Through qualitative and critical analyses of interview and trace data and experiment-based quantitative evaluation of algorithmic systems and datasets, my work contributes to both theoretical underpinnings of marginalized identities in the HCI and CSCW literature while also providing practical and actionable implications for design, governance, fairness, and inclusion in online communities. Inspired by postcolonial computing's capacity to highlight users' agency and generative model of cultural encounter and decolonial focus on rejecting computing's colonial impulse, I view these perspectives as instrumental when examining how colonially marginalized communities across various dimensions express and perform their identities on different online platforms.

## **1.2 Research Overview and Structure of Dissertation**

In this section, first, I will provide a literature review (Chapter 2), a description of the research site (Chapter 3), and a broad-level overview of the methods I used in this research (Chapter 4). Then, Chapters 5-9 present the methodological details, findings, and discussions specific to each empirical study. Finally, Chapters 10 and 11 reflect on the broad implications and conclude the dissertation. A detailed outline of the dissertation is as follows.

### **1.2.1 Chapter 2: Literature Review**

Given the complex constellation of concepts and theories my research builds upon, this chapter will describe central concepts (e.g., identity, colonization) and theoretical frameworks (e.g., postcolonialism, decoloniality) that I draw on throughout the dissertation. I discuss the other related concepts and theories used in specific studies in their corresponding chapters. This section also discusses how colonialism has sociohistorically influenced identity, the objectives of decolonization, how computing technologies reanimate colonial values and structures, and how these technologies support and impede marginalized communities' identity expression and performance.

### **1.2.2 Chapter 3: Research Site and Researcher Positionality**

Given the underrepresentation of Bengali people in computing literature and my interpretivist approach to the studies in this dissertation, this chapter will provide readers with the necessary social, historical, cultural, and political context. It will also explain how the researcher's identity and positionality within the Bengali ethnolinguistic communities as a researcher reflexively shaped the analyses.

### **1.2.3 Chapter 4: Methods**

This dissertation employs a wide range of methods (e.g., trace ethnography, interviews, surveys, NLP, experiments) for different empirical studies. Hence, explaining the methodological details of each study separately in the corresponding chapter improves the readability of the dissertation. Before going into the chapters on empirical studies, this chapter provides a high-level overview of the methods I used and the rationale for choosing certain methods and platforms.

### **1.2.4 Chapter 5: Collaborative Reclamation of Narrative Agency**

Guided by a theoretical framework on processes of decolonization [330], I studied how Bengali Quora users talk about their identity in relation to their colonial past. This work explored how South Asian Bengalis engaged in collaborative identity decolonization work on the platform

to reclaim narrative agency [152]. I found how Bengali users collectively conceptualize the impacts of colonization and mourn through speculative identity play to reaffirm their local and Indigenous ways of being. The study presented how Bengali people of different and opposing political perspectives about the continued transgenerational influences of colonization negotiated their identity to resist and recover from colonial trauma. The findings also unpacked the Bengali people’s dream, commitment, and actionable proposals for reconfiguring regional geopolitical relationships toward pluriversal Bengali sociocultural identity. Outlining the phases of decolonization of Bengali identity, the findings in this study connect with and contribute to the literature on narrative resilience.

### **1.2.5 Chapter 6: Governance Shaping Platform Identity**

In another study on Quora, I demonstrated how a sociotechnical system like Quora reemphasizes colonial values by creating a hierarchy of identities, further deepening division among people through algorithmic coloniality, and imposing identity using collective surveillance and self-imprisonment [149]. Drawing on a conceptual framework that brings together identity performativity, governance, content moderation, and surveillance, this study found that the sociotechnical mechanisms of governance that mediate people’s performances on the platform establish platform identity where in particular identities are privileged while others are pushed to the margins based on dialects, nationalities, and religious affiliations. Whereas the pluriversal perspective emerged as an objective of the decolonization of Bengali identity earlier [152], the emergence of a hierarchy among various Bengali identities by the moderators’ practices hurt the sense of community, online trust, and safety among the users of the platform. In doing so, the platform governance impeded the process of identity expression, performance, and decolonization.

### **1.2.6 Chapter 7: Content Creation to Reimagine Transnational Communities**

A major way in which colonialism has impacted how Bengali people see themselves is through nationalism, where nationalism is often understood through shared language, culture, religion, and geopolitical borders of nation-states. While in the previous study, I explored how the colonially

marginalized Bengali people can engage in collaborative discourse through textual communication to reclaim the narrative agency of conceptualizing themselves on their own terms, in this study, I aimed to understand how multimodal technology can better support decolonial discourses to re-imagine nationalism. Contrary to Quora, where most communication is text-based, YouTube is a multimodal platform comprising videos posted by content creators—individuals or groups and textual interactions among the viewers. To understand this phenomenon, this research draws on a semi-structured interview study with YouTubers who make videos about culturally Bengali people whose lives were upended as a product of colonization and are now dispersed across Bangladesh, India, and Pakistan. The findings of this research describe people’s motivations and strategies for engaging in video-mediated decolonial discourse in transnational contexts. Moreover, this study extends an invitation to foreground the complexities of nationalism in social computing research.

### **1.2.7 Chapter 8: Postcolonial Politics of Content Creation**

Analyzing the same interviews from the previous chapter, I explored how YouTubers face challenges stemming from postcolonial tension in identities and relationships, technological platforms and policies-mediated economic and political power hierarchies, and logistics. In the same chapter, I also examined how YouTubers devise different ways to circumvent these challenges in content creation in postcolonial settings.

### **1.2.8 Chapter 9: Coloniality of Algorithmic Tools and Datasets**

Contrary to the users’ speculation of platform moderators’ identity and positionality in different religious and national communities affecting moderation decisions that I studied in the previous chapter, in this chapter, I investigated the feasibility and potential impacts of automated content moderation instead of human moderators, for governing interaction among colonially marginalized Bengali communities. As natural language processing (NLP) tools, such as the ones for sentiment analysis, are commonly used in automated content moderation frameworks, in this study, I examined how these tools and datasets may be complicit in perpetuating colonial values and bias.

Drawing on identity categories most impacted by colonialism amongst local Bengali communities, I focused my analytic attention on gender, religion, and nationality and developed a Bengali identity bias evaluation dataset. I conducted an algorithmic audit of all Bengali sentiment analysis (BSA) tools available on the Python package index (PyPI) and GitHub, as well as all BSA datasets available through Google Dataset Search and a survey of the developers of these tools and datasets. Despite similar semantic content and structure, my analyses showed that besides inconsistencies in output from different tools, BSA tools exhibit bias based on identity categories and respond differently to different ways of identity expression. Connecting these findings with colonially shaped sociocultural structures of Bengali communities, I explained the downstream implications of these biases.

### **1.2.9 Chapter 10: Toward Decolonization amid the Coloniality of Computing**

Here, I will discuss the broad implications of this dissertation research. First, I will discuss how HCI and social computing researchers can conceptualize colonization as a crisis and its implication for understanding the relationships and collective identities in postcolonial Bengali communities. Second, I will reflect on the methodological challenges, decisions, and ethical considerations in various studies discussed in this dissertation. Third, I elaborate on the implications of HCI and CSCW research, such as postcolonial sociomateriality and engineering activism, with calls for critically studying nationalism and bringing historicist sensibility to cross-cultural computing research.

### **1.2.10 Chapter 11: Conclusion**

In the final chapter, I will summarize the key contributions and future research directions to conclude the dissertation.

## Chapter 2

### Literature Review

#### 2.1 The Relationship among Identity, Marginalization, and Colonialism

Identity is traditionally conceived as one's self-concept—how a person perceives themselves as a physical or social being [193, 234]. In this view, identity relates to how a person sees themselves and how they want others to see them as social and physical beings [221]. It mediates everyday human interaction based on distinguishing people's characteristics, beliefs, and experiences [560]. While identity is often considered an individuated construct, it can also be conceptualized as a collective, where it is shaped by and through people's perceived membership in different social groups [530]. People also construct their identities through broader categorical and/or collective identities, such as those based on race, ethnicity, gender, nationality, sexual orientation, and social class [265, 275, 560].

When connected to broader social and cultural logics [100], collective identity is expressed and experienced through dimensions such as race, ethnicity, religion, gender, sexual orientation, and social class [275]. Various social identities emerge centered around people's perceived membership in different groups [560]. In this inter-categorical [372] view, people's identities are defined across various **dimensions**, such as race, ethnicity, gender, sexual orientation, religion, nationality, and caste. Within each dimension (e.g., religion), people can identify with different **categories** (e.g., Christian) [372]. Importantly, people's identities across various dimensions interconnect and over-

lap, and the consequent intersectional identities collectively shape their unique experiences, social position, and systemic privilege [126, 136]. It is produced and re-produced through ongoing social interactions amongst those who identify with that group or category—it is an ongoing social process [130, 290, 530]. Hence, identities are too complex and fluid to assign to fixed categories—what scholars dub anti-categorical identity [372].

Yet, people often work to assign themselves and others to social groups. They do so through the logic of sorting and categorization, which can perpetuate inequality [79, 177]. Thus, one of the primary ways marginalization happens is based on people’s identity [175, 518]. Marginalization is a process where people in a society are pushed to the fringes and denied their voice or place within it [175, 575]. As people are marginalized based on their individual and/or intersecting identities [192], they experience barriers in living their daily lives, being themselves, and seeing themselves represented as manifest in how marginalization becomes institutionalized and normalized in societal structures and systems [160, 601]. Marginalization of different identities is normalized through cultural hegemony [126, 136]. Cultural hegemony is a system of ideas, practices, and social relationships embedded within private and institutional domains as a mechanism of power and control. Through cultural hegemony, people are categorized as a mechanism of power where some identities are considered “normative” while others are considered non-normative. In other words, people experience everyday harm and are marginalized for being born Black, Queer, or into a lower Caste.

### **2.1.1 How Colonialism Marginalized Collective Identities**

Colonialism, broadly conceived, is the policies, ideologies, and practices through which external powers migrate to and exercise control—full or partial—over a country and its people [325, 352, 546]. Historically, it has served as a primary mechanism for marginalizing people’s identities. Colonialism externally imposed myriad social classification, categorization, and hierarchization schemes in accordance with invented differential ontological densities across regional and global scales [407, 444]. Through colonialism, especially prolonged periods of it, external forces can al-

ter and shift local and indigenous social structures, norms, practices, and economies in profound and long-lasting ways (e.g., shifting how people think about their various and often intersecting national, lingual, and religious identities). This relates to postcolonial theorist Homi Bhabha's [62] concept of hybridity. Since people's identities are not fixed in time and space, hybrid identities are produced through processes whereby colonizing forces establish new logics that merge the language and rituals of "the locals" with that of colonizing forces.

In his formative work, Frantz Fanon [199] wrote extensively about the impacts of colonialism on the collective minds of colonized peoples. Through an exploration of colonized Africa, he draws attention to colonial racism, where colonists often viewed native and Indigenous cultures as inferior to those of the colonial rulers, thus actively working to replace native and Indigenous cultures with those of the colonial power. This process of replacement is dubbed cultural assimilation, which has profoundly impacted those who have been colonized—a trauma felt at both individual and collective levels. First, this assimilation prevents colonized people from developing an independent sense of identity through the destruction and/or theft of cultural sites and artifacts. This "cultural genocide" [577] erases indigenous knowledge and culture [325], denies later generations opportunities for understanding their own cultures. Moreover, this assimilation equates whiteness with superiority and pureness, while blackness is equated with inferiority and evil. Taken together, this process of racial hierarchization leads colonized people to see themselves as subhuman and perpetually experience psychological trauma [198]. This transgenerational psychological submissiveness of colonized people is also conceived as colonial mentality [417]. Colonized peoples suffer from this psychological trauma for generations that perpetually influences how they see themselves, i.e., their identities [198].

Colonial cultural assimilation has shaped people's experiences with racism and colorism (a process that privileges light-skinned people of color over the ones with dark skin [274]) in dramatic and important ways. Today, xenophobic sentiments and heavily militarized borders across the globe are living embodiments of such colonial legacies [89]. Linda Tuhiwai Smith [566] argues that we still live in a world that continues to portray whiteness as an indication of superiority. For

example, cosmetic products advertised through the Internet, television, and print media continue to perpetuate whiteness through the promotion of lighter skin that enforces and re-enforces beauty ideals shaped by whiteness [274]. We also see this in South Asian countries like India and Pakistan, whose inhabitants have an obsession with lighter skin color, showing how the legacy of British colonialism endures [286, 510]. As a result of colorism, India stands as one of the largest consumers of fairness cream products [514]. Further evidence of how colonial mentality continues to shape people's identities is also visible through how the language of colonizers is often used by many people as a preferred mode of communication, even when the majority does not readily understand it. For example, in the Indian subcontinent, the regional languages are Hindi, Bengali, Telegu, Marathi, and more. English was introduced during the British colonial era and is still used as the main language in educational and administrative contexts despite colonial rule ending in 1947. With a low literacy rate (77.7%) [279], many people in a country of 1.35 billion [571] cannot understand and communicate in English. Nonetheless, English remains the dominant language, marginalizing regional languages [300].

While colonization has deeply impacted people's identity, coloniality refers to its enduring and pervasive effects on the local and indigenous communities even after the direct colonial rule has ended [383]. It shapes the hegemonic structures of society by perpetuating colonial impacts on social, economic, political, and cultural dynamics.

### **2.1.2 How Colonialism Impacted Social Identities in Bengali Communities in the Indian Subcontinent**

The British colonial rule in the Indian subcontinent—comprising modern-day Bangladesh, India, and Pakistan—exemplifies the legacy and impacts of colonialism on people's identities. For a broader readership, I will discuss the history of this colonization in detail in Chapter 3. Colonization created myriad fractures in the relationships and diverse national, religious, linguistic, gender, and other identities within the Bengali communities who are native to the Bengal region of the subcontinent [219, 286, 430].

Let's closely look at colonization's impacts on various identity dimensions separately. Among other dimensions of identity, European colonialism imposed its conceptualization of gender on many indigenous communities [355]. Scholars have studied colonized Bengali societies to understand the complex relationship between colonialism and gender [168, 524]. British colonization, they argue, produced a particular kind of masculine identity, wherein the "manly Englishman" was contrasted with the stereotyped "effeminate Bengali" to justify British rule and denigrate Bengali culture [524]. Such a colonial idea of masculinity had profound impacts on gender and ethnic relations. This view led to the stereotyped views of Bengali men in colonial India [168, 449] and the reinforcement of "traditional gender roles" in Bengal [523]. This minimized women's sociopolitical participation and voices [537].

The imposition of European standards also distorted people's religious values and perceptions of the Indian subcontinent. Scholars have attributed the rise of religious extremism and the violence against minorities in the region to colonial values and divide-and-rule practices [156, 406]. They argue that religion-based nationalism is a reactive ideology that emerged in response to the challenges posed by colonialism and the West, where local people have adopted many ideas and practices of Abrahamic religions, such as the emphasis on a single, monolithic God [406, p. 24] and the belief in a chosen people [406, p. 101]. Especially due to cultural assimilation—the idea that colonizers' culture is superior to that of the native communities [199] and cultural genocide—the destruction and theft of cultural sites and artifacts [577], as the colonized subjects were denied the opportunities to explore, understand, and practice their own culture, local and native communities' self-perception regarding religion changed. Moreover, the British colonizers amplified, exploited, and institutionalized local communities' religious differences and divisions [111].

Across the world, colonizers introduced classifications to partition different nation-states based on their own perceptions of nationhood and societal groupings of the native communities (e.g., two-nation theory in India-Pakistan) [243]. Such outlooks disregarded the latter's intricate self-perceptions and interconnectedness [111]. As described earlier, British colonizers partitioned the Indian subcontinent in 1947, prioritizing religion as the only dimension of people's collective

identity. In the context of Bengal, West Bengal, with its upper-caste Hindu majority, was annexed to India, while East Bengal, characterized by a Muslim and underprivileged-caste Hindu majority, became a part of Pakistan [506]. This displaced millions of Bengalis as refugees across the India-Pakistan border [431] and marginalized the Bengali people under Pakistani subjugation [18] as the long geographic distance and myriad cultural differences between Pakistan and East Bengal were overlooked in this colonially imposed idea of nationality. Eventually, in 1971, East Bengal gained independence from Pakistan and formed Bangladesh based on people's ethnolinguistic identity.

Overall, among myriad dimensions of marginalization, colonization crucially impacted the expression of social identities in the context of Bengali communities by impacting their perception of gender roles of men and women, the religious division of Hindus and Muslims, and the socio-economic structures and political consciousness culminating in Bengali communities assuming different nationalities (e.g., Bangladeshi and Indian).

## 2.2 Decolonization as Regaining Control Over Identity

How can people decolonize their identities or work to unpack and make sense of the impacts of colonization on themselves and their social systems? And what should the goals of decolonization work be? To address these questions, I draw on postcolonial and decolonial scholarship.

Postcolonialism emerged as an ideological discipline around the works of diasporic scholars (e.g., Edward Said [473], Gayatri Spivak [537], Homi Bhabha [62], Dipesh Chakrabarty [107]) from the Middle East and South Asia who, for the most part, refer back to those locations and their imperial interlocutors. They look at the European colonialism that took place mainly in the nineteenth and twentieth centuries and its impacts, especially in cultural domains [63]. In contrast, decolonial scholars (e.g., Frantz Fanon [198, 199], Walter D. Mignolo [383, 387], Aimé Césaire [102], Sylvia Wynter [612], Anibal Quijano [444])—often belonging to Latin American and African diaspora [407], explore colonialism from the fifteenth century and onward, articulating a rejection of racial supremacy of the West over its colonial subjects [63]. While often perceived as sepa-

rate academic discourses, in this research, I draw on Bhabra [63] in seeing these perspectives as complementary and fundamental to my exploration as they highlight different experiences with colonialism that are important to reflect on in any exploration of this topic.

With the pervasiveness of colonialism, scholars have studied how colonialism continues to shape societal structures and the ways to resist those influences. More specifically, decolonial scholars explore two primary phenomena—decoloniality and decolonization. Decoloniality focuses on challenging the ways in which coloniality has shaped ontologies and epistemologies [387], while decolonization pays explicit attention to removing colonizers, exploring and rejecting their influence, and shifting towards re-establishing independent nations [198]. While these two phenomena are often presented as separate, they are highly intertwined and can happen simultaneously. According to decolonial scholars [198, 330], true decolonization is achieved through a social, political, and economic reformation that reflects the values and ways of being of the local and indigenous people instead of bearing marks of colonial ideologies. My work explores the relationship between decolonization and identity.

Postcolonial scholars called for challenging the dominant assumptive orientalist framework around identities of different communities [473] and encouraged local and indigenous communities to take ownership of their own narratives [62]. Decolonial scholars have also advocated for similar interrogation, revision, and reformation. Broadly speaking, the process of undoing the transgenerational impact of colonization in colonial and postcolonial societies is called decolonization [198, 330]. Beyond establishing sovereign nation-states, the emergence of which emergence of nation-states has typically been viewed as the end goal of anti-colonialism [198, 330], the primary objectives of this process are reclaiming and reaffirming people's indigenous identities [199] and reforming sociopolitical and economic structures in a way that reflects local values and perspectives [198]. Beyond its material aspects (e.g., land rights) [565], decolonization should also be conceptualized as socio-psychological [199].

As colonial powers created hierarchies and norms around identity (e.g. based on race, gender, sexuality, religion, language, and culture) [330, 550] as part of identity colonization, I am interested

in understanding the process of identity decolonization—the identity work that people engage in to unpack, revise, reaffirm, and strengthen their local and indigenous identities [175, 254, 284].

### 2.3 Technology and Digital Identity

In the context of computing, scholars have explored the ways in which digital platforms also come to exhibit a kind of coloniality. In CSCW and CHI, the formative work of Irani and colleagues [282] developed the conceptual lens of postcolonial computing. In this work, the authors argue that sociotechnical systems are designed with values, and as systems migrate globally, technology can embody a “colonial impulse.” To define “colonial impulse,” identified “a series of considerations” that relies on and reinforces universality, reductionist representation, and colonial hierarchies and politics [180]. Much like how colonists migrated to foreign lands and engaged in practices that have revised people’s cultures and norms, so too can technology. Moreover, postcolonial computing scholars have highlighted how, in the context of design, this kind of work often emerges top-down [282, 381, 440]. This top-down design paradigm has a dramatic effect—it continues to reinforce hegemonic power structures, bias, and norms, which can impact people’s identities and identity expression. When computer systems embody pre-existing biases, they can discriminate against populations often based on identity [215]. Similar to how the identities of Bengali communities across various dimensions such as race, gender, nationality, religion, etc., and their expressions have been impacted by colonialism, in studying the *colonial impulse* of algorithmic construction of digital identities reinforce normative views, and reanimate colonial hierarchies and prejudices by regarding certain identities as more positive or negative.

Scholars have explored how digital platforms afford and enable the opportunity for people to express their identities and build communities for different identities [265, 275]. More broadly, scholars have found that digital platforms, such as Facebook, Reddit, and Archive of Our Own, have supported identity expression and the enactment of social support around people’s identities [187, 252]. The CSCW and CHI community has examined the ways in which digital platforms support

people across a range of life changes, such as when moving to college [137, 529], transitioning out of the military [499, 504], coming out as LGBTQ+ [187, 252, 253], and more. Conversely, scholarship has also highlighted the ways in which digital spaces might threaten people's identities and identity expression [162, 173, 241, 252]. For example, exploring the performance of fatherhood on social media, Ammari and Schoenebeck [25, 27] find that the performances related to sharing information about children or fatherhood were stigmatized. Similarly, the work of Haimson and colleagues [252] illustrates how, during gender transitions, the presence of family and friends on Facebook can serve as both a source of stress and support. Most related to my work is scholarship exploring the relationship between governance, moderation, and decolonization in online spaces. Dosono and colleagues explored the identity work of Asian-American and Pacific Islander (AAPI) communities on Reddit [172, 173, 175]. By studying identity work as a process of deliberation, the authors [173] illustrate how AAPIs use Reddit to push back against the idea of monolithization of the AAPI community as a model minority group. In building on this work, they also examined the ways in which moderators on Reddit engaged in strategies of decolonization [175].

In my work, however, I focus on decoloniality with a "historicist sensibility" as recommended by Soden and colleagues [534], through which we can better understand the historic and complex entanglement of race, ethnicity, gender, religion, nationality, caste, and socioeconomic status in the context of the Global South and colonially marginalized native communities [144] and how the historical incidents and colonial policies shaped nationalism, contemporary social practices, cultural identities, geopolitical relationships, economic systems, and challenges both at individual and collective levels, how people use sociotechnical systems to understand and navigate these complexities, and how coloniality perpetuated by these systems impede such exploration of native and Indigenous people.

## Chapter 3

### Research Site and Researcher Positionality

Scholarship in HCI and CSCW has broadly contributed to my understanding of the relationship between colonialism and technology in myriad ways, starting with formative work in postcolonial computing and decolonial computing [17, 153, 282]. In this research, I focus on decoloniality building on the call to action emphasized by Soden and colleagues [534], who argue that many studies in social computing are missing a “historicist sensibility.” I argue that an interpretivist perspective informed by the regional sociocultural history and guided by decolonial and postcolonial literature can help one better understand the historic and complex entanglement of race, ethnicity, gender, religion, nationality, caste, and socioeconomic status in the context of the Global South and colonially marginalized native communities [144]. Overall, a contextual background about the research site is crucial to understanding how the historical incidents, policies, and perceptions of coloniality shaped people’s self-perception, affected contemporary social practices, cultural and collective identities, the conceptualization of nationalism, geopolitical relationships, economic structures, and embedded colonial structures and values within sociotechnical systems [180, 282, 534]. Moreover, in studying marginalized communities, a researcher’s identity may reflexively bring certain affinities into perspective [343, 491]. Therefore, in this chapter, I will introduce the Bengali people and the language, briefly describe their history with colonization, and discuss my positionality as a researcher within these communities to better contextualize the research presented in this dissertation.

### 3.1 Bengali People and Their History with Colonization

The Bengali (endonym: Bangali, বাংলা, IPA: /ba.ŋa.li/) people are one of the largest ethnolinguistic groups (approximately 259.89 million [5, 456]) of this region, whose native language is Bengali (endonym: Bangla, বাংলা, IPA: /ba.ŋla:/). They are local and indigenous<sup>1</sup> to the Bengal region in the Indian subcontinent.

Scholars have different opinions about the geographic demarcation of the Indian subcontinent [68, 376]. Scholars agree that the region consists “at least of India, Pakistan, and Bangladesh” [82]. However, in that case, the term becomes subject to the politics of nomenclature. While some sources describe the region using the phrase “Greater India” [587], Pakistani national historiography often champions the phrase “Indo-Pakistan subcontinent” [364]. Again, scholars who use the term “South Asia” interchangeably with the term “Indian subcontinent” view Afghanistan, Bhutan, Nepal, Sri Lanka, and the Maldives as parts of this region [414]. However, some scholars have argued against such views. According to them, Afghanistan is a central Asian country [370, 376], and while Sri Lanka and the Maldives have cultural similarities with the countries in the Indian subcontinent, these island countries lack geographic contiguity with the region [376]. In the case of Bhutan and Nepal, these countries used to be British protectorates and not British colonies [422]. Therefore, in my study, I focused on Bangladesh, India, and Pakistan—three countries that, scholars unanimously agree, are parts of the Indian subcontinent and were subjugated together under British colonial rule. India and Pakistan got their independence in 1947. Bangladesh gained independence in 1971 (more on this later in this section). Given the 190-year history of colonialism in the subcontinent, it is not possible to provide a detailed and exhaustive account. Hence, I will provide contextual information to help readers understand the dissertation.

---

<sup>1</sup> Ways to conceptualize indigeneity and identify Indigenous people vary across different regions and are subject to emancipatory politics [54, 195]. For example, in the Indian subcontinent, a region with long history of human migration, though “indigenous” is usually used to identify the tribal *adivasi* groups (e.g., Bhil, Munda, Santhal, etc.) [84], non-tribal groups like the Bengalis, Gujratis, and Oriyas, also have a long history of settlement [613]. To inclusively recognize these communities who lived in the region since long before the British colonization, following prior work [152], I am using the phrase “local and indigenous” communities.

The Mughal Empire is part of the late medieval to early modern history of the Indian subcontinent. In its zenith, it was considered a global leader in the world and its global economy, producing 25% of the world's total industrial output [30, 433]. Bengal, in particular, was the most prosperous Mughal province, generating 50% of the empire's total Gross Domestic Product (GDP) [15, 315, 468]. This prosperity of Bengal continued under the rule of *nawabs* [meaning viceroy] as the Mughal empire's influence declined in the region in the 18th century. During this time, the British East India Company (EIC) arrived in the subcontinent along with other European EICs. The operation of British EIC in Bengal got involved in conflicts with Siraj ud-Daulah, the then nawab of Bengal, on the issues of fortification, taxes, and revenues, which led to the Battle of Plassey in 1757. After the British EIC, under the command of Robert Clive, was victorious in the war, the British EIC assumed control over the political and economic structure of Bengal, which later extended to other parts of the subcontinent. Thus, historically, Bengal was the first region in the Indian subcontinent to be colonized by the British. Since the beginning of British colonial rule in 1757, the local populations organized several revolutions throughout the Indian subcontinent as a means of fighting for freedom from colonial occupation. The most well-organized and coordinated of these revolutions happened on a subcontinental scale during the Sepoy Mutiny in 1857. Though this revolution was not successful in gaining independence, it ended the rule of British EIC, after which the British crown assumed direct administration and control of the subcontinent. After the tremor created by this revolution, movements with well-defined objectives for political/economic autonomy and independence continued to happen.

Throughout the colonial period, Bengal was the site of anticolonial movements (e.g., *Swadeshi*) driven by Bengali nationalism [310]. To diffuse these movements, in 1905, British rulers divided Bengal into two administrative parts, East Bengal and West Bengal, based on the Muslim and Hindu religion of majorities in each of those respective regions. This partition was nullified in 1911 due to vigorous opposition and protests. *Swadeshi* was an example of those movements, often seen as a crucial factor behind the rise of both Indian and Bengali nationalism [45, 477]. During this late phase of British rule, the divide-and-rule policy inflicted communal conflicts and riots in

cities like Kolkata and Noakhali, devoured the social structures of Bengal. Meanwhile, political parties like the Indian National Congress (INC) and the All India Muslim League were established. INC, under the leadership of Mahatma Gandhi, launched a non-cooperation movement and a quit-India movement demanding the end of British rule. As the British crown agreed to the claim of independence after World War II, Muhammad Ali Jinnah proposed a “Two nations theory” that called for separate nation-states for Hindus and Muslims of the Indian subcontinent. Based on this theory, when the British left in 1947, Bengal was used as a site of partition and partitioned with a religion-based border for the second time. West Bengal, which comprised a predominantly upper-caste Hindu majority, became a part of India, while East Bengal, which comprised a predominantly Muslim and lower-caste Hindu (e.g., *namasudra*) majority became a part of Pakistan (renamed East Pakistan) [506].

Since this partition and the formation of Pakistan were based on both political efforts and popular support, some historians characterize this arrangement as a “partnership” [619]. However, the Bengalis in East Pakistan started to recognize the colonial exploitation by West Pakistan (current Pakistan) immediately after Pakistan’s formation. Gradually, many similar historical events (e.g., the language movement) highlighted their cultural differences and West Pakistan’s social, political, and economic injustice to East Pakistan in a stark manner. Under the leadership of Bangabandhu Sheikh Mujibur Rahman, the Bengali people demanded political and economic autonomy. The following days mark West Pakistan’s denial and unwillingness to give in to these demands, denial of the transition of power after Bangabandhu’s electoral victory, and a military genocide in East Pakistan. Therefore, highlighting more on the realities and effects than the initial intent, many scholars argue that West Pakistan’s myriad ways of economic extortion, political suppression, linguistic imposition, sociocultural marginalization, genocide, and rapes were concerted to oppress East Pakistan as a colony [8, 155, 405]. The assertion of Pakistan as a colonial force is not to weigh the impacts of British colonialism similarly or lightly. Upon Bangabandhu’s declaration of independence, the Bengalis in East Pakistan engaged in a liberation war and emerged as sovereign Bangladesh in 1971 [578].

During 1947-1971 and soon after 1971, many Muslim Bengalis migrated to Pakistan for better economic opportunities [598] and many Hindu Bengalis migrated to India fearing religious persecution [128]. The refugee crisis escalated in the region during and for several years after the partition [431] and Bangladesh's liberation war [251]. Because of such power dynamics throughout the history of Bangladesh, India, and Pakistan, their sociopolitical experience with colonization is complex and nuanced, especially in the case of the Bengali people. People in these three countries suffered under British colonization together until 1947. Then, the Bengali people in East Bengal experienced another 25 years (1947-1971) of foreign sociopolitical and economic rule and cultural imposition by Pakistan. Hence, the Bengali people's experiences with colonialism in West Bengal and East Bengal differ. Similarly, while in the case of British rule, people in Bangladesh, India, and Pakistan were colonized subjects, in the case of Pakistani rule, Bangladesh and Pakistan were essentially in a colonized-colonizer relationship (please refer to the explanation above for this characterization and see detailed rationales in the prior works by [8, 509]).

The Bengali people, historically being part of the same ethnolinguistic group, presently live as inhabitants of two independent countries in a postcolonial era. Over time, living in separate countries has weakened the bonds among them [328]. Many modern regional border conflicts in the subcontinent are also directly traced back to the partition. *Chitmohols* (enclaves) in the Bangladesh-India border used to be a topic of tension until 2015 when the two countries exchanged a substantial number of these enclaves for a simpler international border. Moreover, prolonged colonial effects, such as religious nationalism and geopolitical tension (e.g., control over Kashmir and Jammu), adversely impacted the India-Pakistan relationship over the years [492]. While the Bangladesh-India relationship recently has some vital advancements (e.g., resolving border disputes, sharing waters of common rivers), there is still much room for progress [104, 266, 267].

In Bangladesh, 98% of the people speak Bengali as their native language [200], and Bengali is the state language of the country [237]. In India, Bengali is recognized as the state language of West Bengal [238], Assam [278], and Tripura [542], while several other states have substantial Bengali speaking populations [238]. The global Bengali diaspora also has well-established communities in

many other countries. Since the Bengalis are spread across different countries spanning a vast region, where the social practices within this ethnolinguistic group vary significantly, the Bengali language has many different dialects. Noticeable differences in dialects and social practices can be found among the Bengalis in Bangladesh and West Bengal, India (e.g., *Bangal* and *Ghoti*), even within different regions of Bangladesh (e.g., Sylhet and Chittagong) [80].

### 3.2 Researcher Reflexivity

I come from the Bengali ethnolinguistic community in Bangladesh and identify as a cisgender heterosexual man. In Bangladesh, I was born in a religious minority Hindu community. Within that community, I have been marginalized based on my caste identity. My family name Das (দাস, also spelled in English as Dasa) means “slave” or “servant” in Bengali—the non-Aryan aboriginal natives who were categorized as Shudras [306] (also spelled as Sudras) under the Hindu caste system.

To avoid caste-based discrimination where underprivileged castes are “subjected to unspeakable condemnation” in conservative Hindu society [21, 22], my ancestral family like many other scheduled-caste (তফসিলি জাতি: *tafsili jati*) Hindus [506], decided to stay in Muslim-majority then East Pakistan (present-day Bangladesh). During that time, our family became subject to migration, dislocation, and separation during the partition of 1947 and the Liberation War of 1971. Given the colonial history of nation-states and religion in South Asia and recently increasing communal polarization in the region, I, like many others in my community [3, 244, 540], have faced harassment and prejudices from the majority community. Due to my birth religion and caste-based identity, I have been barred from inter-dining, rejected from relationships, subjected to slurs, recommended not to say my last name in public, and told to pursue certain “caste-appropriate” and not “too ambitious” jobs. These experiences with discrimination have shaped my identity as different and not part of the “normative” society. These experiences inspired me to question and study identity.

While getting higher education has helped me be resilient in many instances of social prejudices, I gradually obtained a critical understanding of how academia perpetuates systematic inequal-

ity and hierarchies (e.g., Bengali being the seventh most spoken language and having a comparable number of native speakers to English has little representation and resources in computational linguistics research [301]). In the milieu of Bengali culture, sociopolitical structures, and history, I want to highlight subaltern experiences in academic discourses. While my long-term research objective is to understand the experiences of different under-represented communities (e.g., Bengali people, religious minorities, underprivileged castes) with computing systems and technology, I explored the intersection of identity, coloniality, and technology in my doctoral research.

## Chapter 4

### Methodological Overview

In this chapter, I will provide a high-level overview of the methods used in this dissertation. I investigate how different platforms and sociotechnical systems (e.g., Quora, YouTube, and NLP tools) support and impede the identity expression of local and Indigenous Bengali communities.

#### 4.1 Phase 1: Trace Ethnography of Quora

My exploration started with understanding user interaction on Quora, a question-and-answer (Q&A) platform that facilitates discussions in a Q&A thread format. Quora was established in June 2009. The objective of the platform is “to connect the people who have knowledge to the people who need it, to bring together people with different perspectives so they can understand each other better, and to empower everyone to share their knowledge for the benefit of the rest of the world” [277]. Quora has an embedded social network structure. The users on Quora can follow each other (as social networking sites like Facebook, Twitter, etc.) and participate in Q&A-based discussions in a threaded structure (similar to Reddit, StackExchange, etc.). Quora offers itself in more than twenty languages, including Bengali. This creates greater accessibility for its potential users and opportunities for more contextualized discussions on regional issues [151]. Online platforms that operate using dominant languages (e.g., English and Spanish) of the European colonizers continue to marginalize regional languages and their speakers. Since Bengali Quora (BnQuora)<sup>1</sup>

supports the native language of the Bengali people and was more likely to be accessible to the community irrespective of their foreign language skills, I chose to study Bengali users' interaction on and with the platform. To familiarize readers with the Q&A thread structure and interface of BnQuora, Figure 4.1 shows a screenshot of a BnQuora Q&A thread.



Figure 4.1: A screenshot of a public BnQuora Q&A thread (<https://qr.ae/pGPOSZ>) accessed from the first author's account. The thread currently has 16 answers, only two of them are shown in the image as examples. Text on the right sides of the arrows describe the component (in bold) and translated short text (in italics).

<sup>1</sup> <http://bn.quora.com/>

Taking an ethnographic approach, I wanted to understand how users interact in a platform that prioritizes their native language, especially to discuss their native identities and how the platform’s sociotechnical scaffolds shape their experience in doing so. Besides me, my advisor (who identifies as an Iraqi-American, cisgender, heterosexual man from a minority group within Iraq) and a collaborator (who is a Danish, cisgender, heterosexual man) were involved in this study. We explored how colonially marginalized Bengali communities use BnQuora to reconstruct their local and native identities (Chapter 5), and the ways in which the broader systems of governance on BnQuora shape users’ experience and mediate the interactions between people of different national and religious identities (Chapter 6). We deployed a trace ethnography approach [222, 423], which combines the richness of participant observation with the wealth of data in logs to reconstruct patterns and practices of users in distributed sociotechnical systems. I wanted to understand emergent decolonization strategies and how different governance mechanisms shape people’s digital interactions on the platform. I developed an API called “quoras” [151] (details about this API’s implementation and usage are described in Appendix A), which we used to collect data from BnQuora. More details about the data collection (e.g., list of keywords, sampling) and analyses (e.g., theoretical and conceptual lenses) in these studies are described elaborately in the corresponding chapters.

## 4.2 Phase 2: Interviews with YouTubers

While the trace ethnography on Quora was useful for understanding the broader themes of identity decolonization on online platforms, the lack of explanation of certain statements, underlying motivation, or clarity about the rationale for particular interaction patterns was not available while studying traces. Moreover, as colonialism impacted people’s identities across the world in myriad and complex ways, I wanted to understand the forms and themes of identity decolonization online in greater detail.

Challenging the notion of a monolithic local and Indigenous identity and culture, it is important to understand how people now in different nation-states engage in discourses about their native identity as potentially mediated by and through nationalism. I selected members from Bangladesh, India, and Pakistan due to the multiple waves of foreign subjugation and colonialism that have shaped relationships among people in these countries. Considering the language differences among these countries, I focused on multimodal platforms like YouTube instead of predominantly textual communication-based platforms like Quora. YouTube, with over 2 billion monthly users as of early 2023 [280], is one of, if not the most, popular online media in the world. YouTube has a large user following in India (467 million, the platform's largest user base), Pakistan (71.7 million), and Bangladesh (34.40 million) [313, 544]. Its widespread adoption signifies its capacity to capture diverse perspectives and content. Moreover, the platform has become a significant space for political discussions, especially among marginalized voices, activists, and individuals passionate about national and global issues. Because of its increasing popularity and capacity to provide avenues for sharing diverse communities' ideas and engaging in discourse, I chose YouTube as the site for recruiting video creators.

In two studies, I aim to understand YouTubers' motivations for making videos about colonially marginalized cultures and strategies for video-mediated decolonial discourse (Chapter 7) and how they navigate the challenges they face in this work (Chapter 8). Throughout this exploration, the research examines how these YouTubers envision communities and interpret nationalism, both of which have been influenced by historical experiences of colonialism. I investigated how technological platforms and policies reify and are affected by postcolonial tension and power hierarchies. To answer these questions, I conducted a qualitative study based on semi-structured interviews with YouTubers who regularly create videos centered around Bengali identity, culture, and people. Besides myself and my advisor, a collaborator (who is a cisgender, heterosexual woman from an interfaith minority Jain-Hindu community in India) also contributed by transcribing some interviews in this study. Before beginning the study, we received approval from the institutional review

board (IRB) for all materials and procedures<sup>2</sup>. Chapter 7 provides detailed information about this study’s recruitment, interviews, and data analyses.

### 4.3 Phase 3: Audit of Algorithmic Systems

The last empirical study presented in Chapter 9 is focused on the coloniality of algorithmic aspects of sociotechnical systems. It picks up the conversation on governance and content moderation, which examined the colonial impulses in the development of these large sociotechnical systems and their various components. Prior research has found that in some cases, users trust AI as much as humans for flagging problematic content [398, 556]. To evaluate the feasibility of using algorithmic systems in automated content moderation, I examined the fairness and biases of such AI tools by conducting algorithmic audits of NLP (e.g., sentiment analysis) tools, algorithms, and datasets. I identified these tools and datasets from Python Package Index (PyPI), GitHub, and Google Dataset Search (further details available in Chapter 9). Traditional approaches to algorithmic audits involve querying an algorithm with a wide range of inputs and statistically comparing the corresponding results [382, 557]. This differs from other tests popularly used in computing and HCI literature. For example, unlike other common experiments in HCI, such as A/B tests in which the subject of the study is the users, in algorithmic audit, the subject of study is the system itself [382]. Algorithm audits are also different from other types of system testing due to their broader scope, resulting in systematic evaluations rather than binary pass/fail conclusions for individual test cases. Moreover, audits are purposefully intended to be external evaluations based only on outputs, without insider knowledge of the system being studied [382].

To understand the relationship between the biases of these tools, datasets, and models and the identities of their developers, I conducted a survey with an “Exempt approval” from the IRB at the University of Colorado Boulder. Besides myself and my advisor, the research team for this study

---

<sup>2</sup> We obtained an “expedited approval” for this study from Syracuse University, where I started my doctoral studies. Before I transferred to the University of Colorado Boulder (CU Boulder) with my advisor, all interviews were conducted, transcribed, translated, and unidentified. We contacted the IRB at CU Boulder, and they informed us that as there is no identifiable data for this research at CU Boulder, IRB oversight is not required.

included two cisgender, heterosexual male collaborators (one identified as an Indian Bengali and the other as a White American). I contacted the developers through email and LinkedIn to learn their demographic information, such as gender, religion, and nationalities. To protect the privacy of these developers, we de-identified the tools and datasets and their corresponding developers by assigning an ID to each tool or repository instead of using its URL for identification. Inspired by ethics literature on using internet resources in research that provide methods for obfuscating people’s online identities to protect their anonymity [90, 209], we further obfuscated the tools by describing their implementation and data at a higher level (e.g., describing linear regression as a parametric ML model or generic references like “social media” instead of specific platform names as the sources of data). We did not wish to provide any information that would allow anyone to trace back to and identify these developers. I provide further details about the experiment setup, choice of statistical tests, and comparison metrics in Chapter 9.

## Chapter 5

### Collaborative Reclamation of Narrative Agency

In this chapter, we explore the agency sociotechnical systems afford people in decolonizing their identities through the process of identity reconstruction and reclamation of narrative agency—the ability to tell their own stories, reclaim their histories, and challenge dominant colonial narratives from native perspectives (elaborated in section 5.1). Particularly, we explore the identity decolonization work engaged in by the Bengali people on Bengali Quora (BnQuora) using trace ethnography [222, 423]. We draw on concepts of colonialism, collective identity, narrative agency, postcolonialism, and decolonization to explore how Bengali people reclaim narrative agency over their colonized identities—a process through which they decolonize their collective identities. We find that BnQuora supports identity decolonization work, and drawing on Poka Laenui’s framework elaborating the five phases of decolonization, we highlight the identity decolonization tactics engaged in by users of BnQuora, including (1) rediscovery and recovery, (2) mourning, (3) dreaming, (4) commitment, and (5) action. The conversations people are having on BnQuora are working to revise colonial narratives that have come to shape people’s identities. People actively engage in narrative tactics by developing new narratives and working to regain agency over their identities and histories, which helps them bounce back from the chronic threat and vulnerability of coloniality. We then discuss the relationship between narrative construction and decolonization through the development of a concept we dub narrative resilience.

## 5.1 Background

This study explores the relationship between decolonization and how people collectively reclaim narrative agency. To situate our contribution, we highlight how colonialism took away people’s agency and control over their own identities. To better understand how people work to re-establish agency and control over their identities, we then introduce Poka Laenui’s theoretical framework for decolonization [330]. We find that Laenui’s framework allows for a clear analytical framing for our analysis through the open coding process of our data. Then, we discuss how people draw on sociotechnical systems to reclaim narrative agency as a form of decolonization work.

### 5.1.1 Colonialism as Shaping Narrative Agency: Losing Control Over Identity

Whereas on a micro-scale, coloniality shapes and reshapes how individuals see themselves and others, on a macro scale, coloniality takes away the agency and power local and Indigenous populations have over themselves and their own identities. For example, historically, Europe has viewed the East through a fabricated concept called “the orient” as the unfamiliar “other,” an irrational, illogical, unknown, uncivilized, and exotic entity [473]. Orientalism became the basis of the depiction of the East [473].

As colonial powers shaped and framed global narratives about local identities, local peoples did not have agency in framing their own narrative—how outsider colonizers perceived them. This phenomenon is also known as narrative agency [246, 537]. In conceptualizing narrative agency as the relationship between power and agency over one’s self and identity, Gayatri Spivak [537] and Ranajit Guha [246] popularized<sup>1</sup> the concept of the subaltern. Through their explorations of the impacts of colonialism on the Indian subcontinent, the authors defined the subaltern as a negative space wherein colonized populations are excluded from the hierarchy of society by and through those in power. It is not just that the colonizers are ignoring the voices of the colonized; rather, they also purposely subvert the will of the colonized to align with the interests of the colonizers

---

<sup>1</sup> The concept was first coined by Antonio Gramsci [240].

as a way of maintaining power and control. Through this process, members of the subaltern are denied agency and voice in shaping their own identities and societies. [246, 537].

By not giving voice or agency to those they were colonizing, colonial rulers justified their material domination and subjugation of the people [63]. Colonial enterprises were mediated by and through a false cultural superiority and a white savior ideology [199, 272]. While colonial powers worked to erase the identities of local populations, postcolonial and decolonial scholars alike have highlighted the fallacy of cultural superiority and white savior ideological perspectives. The work of postcolonial scholar Homi Bhabha [62] introduced the concept of cultural hybridity, arguing that culture is not a static entity and not an essence that can be fixed in time or space. Through this lens, there is no pure Africanness, Britishness, or Indianness, and thus, the binary notion of superior Britishness or inferior Indianness does not hold weight. There is no uncontaminated form of culture. All cultures are identified by a notion of hybridity, and thus, the concept of cultural superiority is invalid.

Through the often invisible reshaping of social structures and norms, scholars draw attention to how colonization created myriad fractures and fissures with the sociopolitical fabric of local contexts, creating conflicts of identity across national, religious, linguistic, and other identities of those who were colonized as a way in which to establish power and dominance [102, 385, 387]. Aimé Césaire points out that one of the main driving forces behind colonialism was market interest—whether the logic of capital prevails [102]. Historically, the primary goals of colonialism included gaining economic resources, increasing political control and power through territorial acquisition, and installing new governments and ideologies to maintain this control [1, 433]. As a way in which to achieve these goals, colonizers dehumanized local and indigenous people’s identities, reducing them to a state of commodity or thing—the process of “thingification” [102]—so that they could be enslaved and exploited for production.

Decolonial scholars have also illustrated how colonizers have exploited indigenous populations through explicit efforts to have them work and act as intermediaries who re-enforce colonial logics. In the context of the Indian subcontinent, the political and economic elites and upper castes

served as interlocutors [31] to support and propagate colonialism, which was yet another way they were exercising power and dominance. Oftentimes, the colonizers created policies that served to enforce specific behaviors among those they were colonizing [143]. Divide-and-rule is an example of such policies. Using this policy, colonial rulers exacerbated the adversarial logics mediating the relationships among the regional populations they sought to control. They made communities in the subcontinent think of other communities as their adversary while exploiting them politically and economically. Colonial rulers highlighted religious differences amongst Hindus and Muslims, creating division which gradually led to the creation of separate communities for each religion, weakening the regional social structure, and eventually the partition of the subcontinent [245, 311, 446, 545]. With the creation of national borders based on religious partitions in 1947, families became separated based on these colonially constructed categories [112], leading to one of the largest refugee crises in history [431]. Today, in modern-day nation states, the contestation of religious minorities' identity by the religious majorities [220, 288, 379, 445] has led to religious persecution, forced migration, and forced conversion [595], and a rise in the politicization of religious identities [77, 293, 441].

The impacts of colonialism continue to dominate how many non-Western societies operate and in shaping people's narrative agency or lack of narrative agency. In the context of computing, scholarship in postcolonial design [180, 282, 381] illustrates how sociotechnical systems also have colonialist tendencies. The formative work of Irani and colleagues [282] demonstrates how, by virtue of being designed in Western contexts and with Western values, as sociotechnical systems migrate and travel to other, especially non-Western contexts, they reanimate colonialism. In this way, sociotechnical systems can continue to perpetuate power and control over others. As such, postcolonial computing discourse is centered around questions of power, authority, participation, and intelligibility with respect to technology design and use in cultural contexts [282]. Yet, scholars have also critiqued the work of postcolonial computing scholars [16, 17], highlighting how by conceptualizing this phenomenon as postcolonial, this assumes that we have moved past colonialism. This is best illustrated by Linda Tuhiwai Smith, who argues: "naming the world as "post-colonial" is,

from indigenous perspectives, to name colonialism as finished business...There is rather compelling evidence that in fact this has not occurred” [566]. Put another way, the impacts of colonialism persist to this day and continue to exert power and control over people’s narrative agency in myriad ways.

### **5.1.2 Decolonization as Re-Establishing Narrative Agency**

In this work, we embrace the perspectives of decoloniality to understand how people are working to reclaim narrative agency, i.e., control over their collective identities. To highlight how decolonization happens, we draw upon Poka Laenui’s work [330], who highlights the process of decolonization and embodies the perspectives of other postcolonial and decolonial scholars like Bhaba and Fanon [62, 198]. He views decolonization as a broader process whereby colonization only ends when colonized people can reframe and revise their colonial mindsets [330]. He articulates this process through five distinct phases, which include rediscovery and recovery, mourning, dreaming, commitment, and action.

In the first phase—rediscovery and recovery—people who have experienced colonization suffer from the concepts of inferiority. This phase drives one to rediscover their history and recover their culture, language, and identity, and thus, is fundamental to the process of decolonization. The second phase, mourning, is a period where a colonized population laments their victimization. Considering colonialism as a long-term trauma, mourning becomes an essential phase of healing. In fact, the mourning stage can accelerate the earlier stage of rediscovery and recovery. Laenui states that the third phase of this process, dreaming, is the most crucial for decolonization, where the broader picture of possibilities is constructed, expressed, and considered for a new social, cultural, and political order through debate, consultation, and discourse. After exploring the panorama of possibilities in the dreaming phase, in the fourth stage, known as commitment, people have to be committed to moving in a direction that culminates in people combining their voices in a clear statement of their desired sociopolitical structure. The fifth and final stage of the decolonization process is called action. This can incorporate a wide range of activities - from a call to reason

to armed rebellion. While many countries achieved independence through armed liberation wars (e.g., Bangladesh) or at least made it a part of their independence movement (e.g., India), the decolonization environment has drastically changed in the last several decades. Though these phases are described as sequential processes, these phases can overlap and occur simultaneously [330].

### 5.1.3 Reclaiming Narrative Agency with Technology

In this dissertation, we are especially interested in how people use sociotechnical systems to decolonize their identities and thus reclaim narrative agency over how they see themselves and others, drawing on the framework elaborated by Laenui [330]. In our work, we focus explicitly on decolonization work as a collective social process, exploring the discourses engaged in by people with colonized histories, echoing how decolonial scholars emphasize the critical role of conversation in the process of decolonization [330, 386].

Studies exploring the experiences of people with identity-based trauma find that for people to make sense of uncertainty, they often communicate with others who have had similar or shared experiences—a process known as collaborative sensemaking [597]. Identity work and expression can also take on the form of identity play, whereby people playfully take on and assume different identities as a mechanism through which to process their own [275]. When examining the uses of sociotechnical systems for identity work and expression as part of the recovery from traumatic ongoing experiences, HCI and CSCW scholarship has focused on people’s uses of online communities and social media during life changes, such as residential moves [516], the transition from high school to college [399], recovering from experiences with domestic abuse [122], coming out as LGBTQ [187], transitioning out of the military [504], gender transition [252], and job loss [96]. Prior work has underscored the important role sociotechnical systems play in helping people manage life change, especially through the opportunities afforded in seeking social support and developing communities.

Although prior work has highlighted how sociotechnical systems support identity work and identity expression, sociotechnical systems can also further marginalize people’s identities or complicate identity work. Haimson and colleagues [252], through an exploration of Facebook use during

gender transitions, found the platform to be both a source of stress and support. In a study exploring the uses of Facebook among college students from disadvantaged backgrounds, Morioka and colleagues [399] highlighted how the platform made it difficult for students to identify supportive mentors to help them navigate their transition to and through college. Similarly, in their study exploring fathers' experiences sharing photos and other information about their children, Ammari and Schoenebeck [25] found that fathers experienced stigma in expressing this part of their identities online. More recently, in exploring the ways in which algorithmic systems like TikTok shape identity work and expression, scholars have found that algorithms perpetuate harmful societal norms and marginalize non-normative identities [309, 518]. Taken together, these systems can take away people's narrative agency.

The ways in which sociotechnical systems can limit or impede people's narrative agency can be related to the colonial impulse inherent within technology [282]. As previously described, the work of Irani and colleagues [282] highlighted how technology is designed in the West and with Western values, and when it migrates elsewhere and encounters different cultures, it can perpetuate coloniality. In building on this formative work, HCI and CSCW scholars have adopted both decolonial and postcolonial perspectives to empirically study the relationship between technology and local and indigenous practices [149, 175, 461, 594]. Building upon Irani and colleagues' work on postcolonial computing [282], researchers demonstrate how unique local practices emerge around technologies that promote sustainability [158, 339, 461] and how technologies work as a medium of imposition of Western standards on the Global South [392, 451]. Moreover, decolonial computing scholars have studied the innovation practices in the Global South that reflect local needs and values [69, 110]. Besides empirical studies, efforts have been made within the HCI community to outline an agenda for decolonizing the pedagogies of the discipline [19, 337].

Influential work by Dourish and Mainwaring argues that both colonial narratives and designs of ubiquitous computing systems subscribe to a notion of universality and share a commitment to reductionist representation for the purpose of comparison, evaluation, understanding, and prediction [180]. HCI studies from postcolonial perspectives have continued to scrutinize the utopian

promise of development through technology and often found those to be unsuccessful in solving complex social problems with overly simplistic solutions [440, 594]. Still, many modern sociotechnical systems (e.g., predictive policing, AI datasets) are built upon these ideologies [397], which conceals the social reality of the people it describes and hides the positionality of those who employ those [56]. Recent scholarship has started to examine how racial identities are constructed within sociotechnical systems and how the politics of representation exhibits coloniality [486, 489].

Prior work has focused mostly on how sociotechnical systems are inherently colonial and ways to combat this. Our work, however, builds on prior work and explores its foil—how sociotechnical systems afford people collaborative opportunities for engaging in the work of identity decolonization and thus reclaiming narrative agency. In this chapter, we explore how people are using the Quora platform to reclaim narrative agency. By understanding how local practices emerge within a colonially marginalized ethnolinguistic group around an online platform and studying how the community rejects colonial influence on its identity using technology, our work contributes to the postcolonial and decolonial conversations as well as the HCI literature on marginalized identities.

## 5.2 Methodological Details

This section will provide details about trace data collection (e.g., preparing a list of search phrases and keywords), explain our data analysis approach, and reflect on this study’s limitations.

### 5.2.1 Data Collection

This study draws on a dataset comprising Q&A threads from the BnQuora platform using online data collection beginning on 15 May 2020 and ending on 15 July 2020. Importantly, the data presented in this study contains Q&A threads dating back to the launch of the platform through the end date of our data collection process. We collected data using the quoras API [151] combining purposive sampling [555] and snowball sampling [236].

As an initial step in our purposive sampling [555] process, the authors engaged in in-depth discussions to generate an initial list of keywords/phrases. We emphasize that due to the long

history and far-reaching impacts of colonialism experienced by the Bengali people, it is difficult to create an exhaustive list of search terms. In preparing the list of keywords and phrases, we focused on (a) concepts related to colonial history and Bengali identity (e.g., উপনিবেশবাদ, বাঙালি সংস্কৃতি: colonialism, Bengali culture); (b) the sociopolitical and cultural history of Bangladesh, India, and Pakistan such as important historical figures (e.g., ব্রিটিশ ইস্ট ইন্ডিয়া কোম্পানি: British East India Company); (c) places (e.g., ঢাকা, কলকাতা: Dhaka, Kolkata); (d) historical events with their years (বঙ্গভঙ্গ - ১৯০৫ partition of Bengal - 1905); and (e) emergent political ideology (e.g., স্বদেশি: Swadeshi ). While preparing this list, we were also aware of and used both archaic and revised spellings of the keywords (e.g., বাঙালি and বাঙালী are different spellings to mean “Bengali people”) and widely used synonyms and similar phrases (ভারত, ভারতবর্ষ, and ইন্ডিয়া all mean “India”, where the first two words are endonyms while the last is the exonym for the country). The list of keywords and phrases is presented in Table 5.1 with translations and/or additional explanations about the keywords.

Table 5.1: List of keywords and phrases. The search keywords/phrases are sorted in the alphabetic order of the second column for easier lookup for non-Bengali speaking readers.

Keyword/Phrase	Translation and Explanation (if needed)
১৭৫৭, ১৮৫৭, ১৯০৫, ১৯১১, ১৯৪৭, ১৯৫২, ১৯৬৬, ১৯৭১	1757 (Battle of Plassey), 1857 (Sepoy mutiny), 1905 (Partition of Bengal), 1911 (Nullification of the partition of Bengal), 1947 (Independence of India and Pakistan; partition of Indian subcontinent), 1952 (Bengali language movement), 1966 (Six-point movement), 1971 (Independence of Bangladesh) (years of important historic events)

**Table 5.1 continued:** List of keywords and phrases

Keyword/Phrase	Translation and Explanation (if needed)
৩৭০ ধারা, CAA, NRC	Article 370 (an article of Indian constitution that gave Jammu and Kashmir special status, recent withdrawal of which has led to unrest and military conflict in the India-Pakistan border [409]), Citizenship Amendment Act, National Register of Citizens (some recent events that are traced back to the British colonial era and have implications on modern days politics [410, 596].)
অসহযোগ, অসহযোগ আন্দোলন	Asahajog movement (non-cooperation of the Indigenous population with the British rulers)
আওয়ামী লিগ/আওয়ামী লীগ, ভারতীয় জাতীয় কংগ্রেস, যুক্তফ্রন্ট, নিখিল ভারত মুসলিম লিগ/নিখিল ভারত মুসলিম লীগ	Awami League, Indian National Congress, United Front, All India Muslim League. These are the prominent political parties during the liberation movements of Bangladesh, India, and Pakistan.
বঙ্গবন্ধু/শেখ মুজিবুর রহমান/শেখ মুজিব	Bangabandhu Sheikh Mujibur Rahman (prominent leader and the father of the nation of Bangladesh)
বাংলাদেশ, ভারত, পাকিস্তান, ঢাকা, চট্টগ্রাম, সিলেট, দিল্লী/দিল্লি, মুর্শিদাবাদ, কোলকাতা/কলকাতা, আসাম, কাশ্মীর, লাহোর, করাচি, পাঞ্জাব	Bangladesh, India, Pakistan, Dhaka, Chittagong, Sylhet, Delhi, Murshidabad, Kolkata, Assam, Kashmir, Lahore, Karachi, Punjab (the countries in the Indian subcontinent and cities of historic importance in the colonial period)

Table 5.1 continued: List of keywords and phrases

Keyword/Phrase	Translation and Explanation (if needed)
বাংলাদেশ ভারত সম্পর্ক, ভারত পাকিস্তান সম্পর্ক, বাংলাদেশ পাকিস্তান সম্পর্ক	Bangladesh-India relationship, India-Pakistan relationship, Bangladesh-Pakistan relationship
পলাশীর যুদ্ধ/পলাশির যুদ্ধ	Battle of Plassey (the battle where the British East India Company won and started the colonial era)
বাঙালি-সংস্কৃতি/বাঙালিয়ানা	Bengali culture
বাংলা/বঙ্গ/বাঙলা	Bengali language or Bengal region
বাঙালি-জাতীয়তাবাদ	Bengali nationalism
বাঙালি/বাঙালী/বাঙালি-জাতি	Bengali people
ইস্ট ইন্ডিয়া কোম্পানি/ব্রিটিশ ইস্ট ইন্ডিয়া কোম্পানি	British East India Company (a British trading company that started the colonial domination in the Indian subcontinent)
ব্রিটিশ ভারত/ব্রিটিশ ইন্ডিয়া	British India
ব্রিটিশ সাম্রাজ্য	British imperialism
ব্রিটিশ শাসন, ব্রিটিশ রাজ	British rule, British Raj (Specifically the rule of British crown of the Indian subcontinent from 1857 to 1947)
ঔপনিবেশিক	Colonial
উপনিবেশবাদ	Colonialism
উপনিবেশ	Colony
পূর্ববঙ্গ, পূর্ববাংলা, পূর্ব বঙ্গ, পূর্ব বাংলা	East Bengal (the name of present day Bangladesh during the British period.)
পূর্ব পাকিস্তান	East Pakistan (the name of present-day Bangladesh during the Pakistan period)

Table 5.1 continued: List of keywords and phrases

Keyword/Phrase	Translation and Explanation (if needed)
ঘটি ও বাঙাল	Ghoti (Bengali people from West Bengal) and Bangal (Bengali people in Bangladesh)
বৃহত্তর-বাংলা	Greater Bengal (comprising of present-day Bangladesh and the West Bengal state of India)
বাংলাদেশের স্বাধীনতা, ভারতের স্বাধীনতা, পাকিস্তানের স্বাধীনতা	Independence of Bangladesh, Independence of India, Independence of Pakistan
ভারতীয় উপমহাদেশ	Indian subcontinent
ভাষা আন্দোলন	Language movement (demand for recognizing Bengali as a state language of during the Pakistan period)
স্বাধীনতা সংগ্রাম, স্বাধীনতা যুদ্ধ/মুক্তিযুদ্ধ	Liberation movement (usually used in the context of India and Pakistan), Liberation war (usually used in the context of Bangladesh.)
গান্ধি/গান্ধী/মহাত্মা গান্ধি/মহাত্মা গান্ধী	Mahatma Gandhi (prominent leader and the father of the nation of India)
মুঘল সাম্রাজ্য, সুলতানী, নবাবী	Mughal empire, Sultani, Nawabi (the governments in different parts of the Indian subcontinent before the arrival of the British colonizers.)
মুহাম্মদ আলী জিন্নাহ/জিন্নাহ	Muhammad Ali Jinnah (prominent leader and the father of the nation of Pakistan)
বঙ্গভঙ্গ/বঙ্গ ভঙ্গ, বঙ্গভঙ্গ রদ	Partition of Bengal, Nullification of the partition of Bengal

Table 5.1 continued: List of keywords and phrases

Keyword/Phrase	Translation and Explanation (if needed)
দেশভাগ, ভারত ভাগ, পার্টিশন	Partition of country (usually refers to the partition of 1947, partition of the Indian subcontinent)
ভারত ছাড়/ভারত ছাড়ো/ভারত ছাড় আন্দোলন/ভারত ছাড়ো আন্দোলন	Quit-India movement
উদ্বাস্তু/রিফিউজি	Refugee
ক্লাইভ/লর্ড ক্লাইভ/রবার্ট ক্লাইভ	Robert Clive (the first British Governor of the Bengal Presidency)
সাম্প্রদায়িকতা, সাম্প্রদায়িক দাঙ্গা	Sectarianism, Communal riots
সিপাহি বিদ্রোহ/সিপাহী বিদ্রোহ	Sepoy mutiny (the first organized revolt against the British colonial rule)
সিরাজ-উদ-দৌলা/সিরাজ উদ দৌলা/সিরাজউদ্দৌলা	Siraj ud-Daulah (the last independent <i>Nawab</i> of Bengal)
৬-দফা/ছয়-দফা/৬ দফা/ছয় দফা/৬ দফা আন্দোলন/ছয় দফা আন্দোলন	Six-point movement (demand for autonomy of East Pakistan)
স্বদেশি/স্বদেশী, স্বদেশি আন্দোলন/স্বদেশী আন্দোলন	Swadeshi movement (subcontinental indigenous people's economic uninvolvement with the British rulers)
দ্বিজাতি তত্ত্ব	Two nations theory (a proposal that identifies Indian Hindus and Muslims as two distinct nationalities and calls for two separate countries for them.)
পশ্চিম পাকিস্তান	West Pakistan (the name of present-day Pakistan during 1947-1971)

Though the quoras API [151] supports keyword/phrase searching directly, it discourages the use of this function. To abide by the advisory *robots.txt* file of the BnQuora platform, we searched for Q&A threads containing the search keywords or phrases directly using a web browser.

Using this sampling approach, we collected 1388 unique Q&A thread URLs. However, because of the myriad contexts in which some of the keywords can be used, all retrieved Q&A threads were not relevant to our study. For example, let us consider the following two Q&A threads’ titles retrieved by searching with the keyword **ঔপনিবেশিক** (meaning “colonial”): *“How do you see the tendency to speak English more than necessary – the evils of two centuries of colonial slavery or the pattern of disrespect for one’s own culture?”* and *“If you had the opportunity to make the first colonial trip to the Mars, would you go?”*. Here, the first question is relevant to our study, whereas the second one is not. To determine which Q&A threads were relevant to the study, i.e., whether the thread is related to colonial history and its impacts on Bengali people, the first author and an undergraduate student—both of whom are born and raised in Bangladesh, separately labeled the relevance of the collected Q&A threads based on the original Bengali texts in the questions. Only the Q&A threads labeled relevant by both were retained in the keyword-based dataset, yielding 693 Q&A threads through purposive sampling.

To increase the breadth and volume of the dataset used for analysis, we also included the related Q&A threads recommended by the Quora Recommendation Algorithm (QRA) [614], as a form of snowball sampling [236]. The QRA trains itself to learn users’ interests and similarities among topics and Q&A threads, and thus, including its suggestions, diversifies and broadens the scope of our dataset. For each of the Q&A threads in the keywords-based dataset, we retrieved the QRA suggestions about related questions. After retrieving 2172 recommended Q&A threads and employing the same relevance labeling approach as previously described, 778 Q&A threads were retained in the recommendation-based dataset after relevance labeling. Including the suggested Q&A threads recommended by QRA led to the discovery of data on topics that were absent from our purposive sample. Some examples of such topics captured through snowball sampling are: **হিন্দুস্তান**: Hindustan (a colloquial endonym of India), **সত্যগ্রহ**: Satyagraha (literal translation:

“holding firmly to truth” - a particular form of nonviolent civil resistance), ভারতের-মুসলিম-শাসন: Muslim rule in India, etc.

The keywords-based dataset and the QRA recommendation-based dataset were merged to create a final dataset comprising 1471 Q&A threads. In following the terms and conditions of Quora, we cannot make our dataset public<sup>2</sup>. However, the dataset could be replicated if our data collection strategy were followed.

### 5.2.2 Data Analysis

We used a dual inductive and deductive approach. The data collected from BnQuora was consolidated and subjected to open coding and inductive analysis [380]. We used the qualitative data analysis software Quirkos<sup>3</sup> for the coding process. The first author conducted the preliminary coding of the data and met the second author twice every week with English translations<sup>4</sup> of several exemplar quotations for each emergent code. We identified the abstract representations of the entities, concepts, and interactions that repeatedly appeared in the Q&A threads. Examples of some inductive themes that emerged are: “defining colonialism-related concepts”, “marginalization of regional culture and knowledge”, “disruption of social structure”. Based on the patterns that emerged, we noted the representations of the entities related to coloniality, concepts under decolonial and postcolonial studies, and interactions that depict different phenomena studied in the aforementioned fields, repeatedly appearing in the Q&A threads.

After completion of our inductive approach, to understand these user interactions on BnQuora better, we engaged in a review of theoretical literature, which brought us to Poka Laenui’s articulation of “Processes of Decolonization” [330]. We found this framework to be particularly productive because of its phenomenological grounding in both history and lived experiences, as well as its capacity to capture the underlying thrust of decolonial literature. Moreover, the phases

---

<sup>2</sup> <https://www.quora.com/about/tos>

<sup>3</sup> <https://www.quirkos.com/>

<sup>4</sup> In the results section of the chapter, we include the English translations of the users’ Bengali quotations only. However, we will be happy to make the Bengali quotations and the URLs of the corresponding Bengali Q&A threads available on request.

of decolonization identified by Poka Laenui are broad and flexible so that these could incorporate the multiple ways that people identified the colonial influences on their identity and pursued the process of decolonization of their identity. As we organized our findings using this framework, we found it comprehensive for our dataset (e.g., we did not need to adjust the framework to account for our findings). The phases were useful in characterizing the act of rediscovering and recovering the Bengali users' local and indigenous identity, mourning their colonial past, and dreaming of their decolonial reconfiguration. Poka Laenui's phases were also helpful in explaining how the Bengali users on BnQuora can commit to their decolonial vision and take action toward repairing their identity. Such a combination of iterative coding and theoretical frameworks is common within the HCI community [329, 504]. In addition to providing an empirical account of the decolonization process of the understudied Bengali population on online platforms (e.g., BnQuora), we emphasize the importance of identity studies in the context of this population and introduce Poka Laenui's take on the process of decolonization which to our knowledge has not been explored and utilized in CHI and CSCW community. While our work stayed true to ethnographic tradition, the context and focus on conversations about colonial histories are inherently political. Moreover, as an important part of the ethnographic tradition, the unique identities and reflexivity of the authors have shaped the interpretations of the data presented in this chapter.

### 5.2.3 Limitations

In this work, we explored the decolonization strategies as a collective identity reconstruction process utilizing the scaffolds of a regional language-based sociotechnical platform. Given the large majority of the Bengali people live in Bangladesh and India, most of the users on BnQuora participate from these countries, with a few users joining from Pakistan. Therefore, due to the user demographics of the platform, the study draws heavily on the experiences of Bangladeshi Bengalis and Indian Bengalis and captures the experiences of only a few Pakistani Bengali users.

### 5.3 Results

On BnQuora, we find that emergent questions and conversations illustrated that people were using the platform to engage in tactics for identity decolonization work. Importantly, these manifested as labor in service of reclaiming narrative agency. Through our analysis of conversational threads, we describe these strategies across the phases of decolonization as outlined by Poka Laenui [330]: rediscovery and recovery, mourning, dreaming, commitment, and action. We emphasize that these phases of decolonization can be experienced at the same time or in various combinations [330].

#### 5.3.1 Rediscovery and Recovery: Collaborative Conceptualization and Healing from Collective Colonial Trauma

The process of rediscovery and recovery—wherein people who have experienced colonialism work to rediscover their history, culture, language, and identity—is the first step towards decolonization [330]. During the colonial period, colonized populations were heavily influenced by a “white savior ideology” that justified the actions of their colonizers [272]. Through the process of colonial cultural assimilation, colonized peoples are made to see themselves as subhuman and, thus, develop a colonial mentality [198, 199]. Through our analysis, we find that BnQuora users developed two tactics for rediscovering and recovering their identities, namely: (1) Collectively Conceptualizing Colonialism and (2) Drawing on Local Perspectives to Challenge Trans-generational Psychological Domination. These tactics served to empower users of BnQuora to make sense of and collectively reconstruct their colonized histories on their own terms.

##### 5.3.1.1 Collectively Conceptualizing Colonialism as Rediscovery: Narrative Authority of the Subaltern

Colonialism has long denied the Bengali people the agency to generate discourse, putting them in the subaltern space—a negative space where colonized populations are excluded from the

hierarchy of society by those in power and thus do not have agency over their own voices and identities and are often forced to align to the interests of their colonizers [246, 537]. On BnQuora, people engaged in collective discourses to conceptualize and define colonialism in their own terms—to reclaim agency, voice, and their own identities. Here, people were dualistically defining colonialism and making sense of the impact of colonialism on their identities. By conceptualizing and coming to terms with colonialism, they were also grappling with how colonialism had impacted their identities, serving as a point through which they could begin to rediscover their local and indigenous selves and re-establish narrative agency in developing their own hybrid identities [62]. Importantly, users are not trying to return to any imaginary pre-colonized pristine Bengali culture; rather, they are working to identify the existence of colonial traces on their lived experiences. In this process, the users attempt to interrupt the Western discourse of modernity by becoming active interrogative subalterns. They reflexively conceptualize critical theoretical perspectives—a crucial step towards decolonization.

This initial work to conceptualize the concept of colonization becomes the starting point through which collective decolonization work happens on BnQuora. Participants sought to conceptualize colonialism and related concepts, such as: “colonialism”, “colony,” “imperialism,” and “neo-colonialism,” to name a few. The following example best illustrates how users were working to understand the term “colonialism”:

*Q: What is meant by colonial rule?*

*P1: ... Colonial rule, according to popular belief and based on various historical events, is a system of government that is usually established by a country or a nation influencing another country or nation through war or commercial tactics, religious views, or propaganda.*

*P2: Colonialism, in the simplest sense of the word, is an “economic-political-social” way of influencing literally and physically. ... Such as British colonialism. ...*

In defining the concept, they used their definition to then reflect on the impacts of colonization on their identities. While the conversations were diverse, many of the conceptual conversations focused on how the practices of colonialism have influenced and restructured the identities of the local populations of the Indian subcontinent. For example, one of the dominant and unfortunate practices of colonialism has been the erasure and/or marginalization of local and indigenous cultures, knowledge, and practices [325, 566]. Users of BnQuora conceptually grappled with how the marginalization of their local practices has impacted their identities. For example, in the following conversation, users discuss the ways in which colonially imposed educational systems and policies destroyed subcontinental knowledge and communal harmony:

***Q:** What is imperialism? How did imperialism affect India?*

***P1:** ... The ugliest British policy is “Divide and rule”. The British have injected this poison into the pores of the people of India. ... The British came to do business in this country and sowed the seeds of a high-low, rich-poor, black-and-white attitude in everything. ... The most important philosophy of India is to achieve spiritual consciousness. ... The widespread use of English mediums at all levels is the cause of spiritual loss for the people of this area. To this day, [these have] left Indians extremely hesitant/divided.*

The conversations on BnQuora served as important points of reflection for people. Considering that they were relegated to the subaltern and colonial mentality had become so deeply embedded into people’s everyday lived experiences, and their identities, users of the platform began to reflect on and re-interpret their past lived experiences. They did so by drawing on the conceptual lenses the community was generating together. By drawing on these lenses, people were beginning to distinguish between what aspects of their lived experiences were influenced by colonial histories versus what was not. For example, one user of BnQuora describes an incident where a person physically hurt another for addressing her using Bengali greetings instead of English ones. In reflecting on this

experience, the user was applying the knowledge generated through their interactions on BnQuora in understanding the colonial influence on perspectives:

*Q: A nurse called a female doctor apa (meaning sister in English). That's why the doctor slapped the nurse. Do you think addressing someone as "apa" is less respectable than addressing as "madam"? Do the colonial influences remain in the minds of the people?*

*P1: ... When starting to teach in our class, a mathematics teacher in our college [college name] told us to respond in English by saying "yes, sir" instead of using Bengali phrases like Uposthit (meaning present in English) while taking attendance record. Many people find pleasure in using English words while speaking. If this is not a colonial effect, then it's mental decay.*

What we can see here is an aspiration among the previously colonized people to behave like their colonizers. To the doctor and the teacher in the shared stories in the thread, to be respected is to be like the English. They equate linguistic practice with the notion of respect while thinking of their native language, Bengali, to be inferior to the language of their colonizers. Through this exercise of sharing stories, the users are locating the existence of collective colonial trauma [198].

As is visible in both stories we have featured above, users of BnQuora identify British colonial influence as a possible reason behind the preference toward the English language over their native tongue. Moreover, we see a certain power distance in doctor-nurse and teacher-student relations where the choice of language is mediated through authoritative decisions and not through individual preferences. These conversations also highlight how users resist Western colonial preferences regarding normative modes of communication. Thus, in effect, BnQuora users are highlighting the dynamics of postcolonial interculturality as outlined by Irani and Dourish [283]. They are not against mixing languages; rather, as elaborated in the following Chapter 6, they celebrate linguistic hybridity [149] and are actively interrogating whether their linguistic practices result from their self-determined preference or external imposition. Through their discussions, BnQuora users

identify, question, and reject colonial superiority in different facets of their lived experiences (e.g., linguistic practices).

In sum, conceptualizing colonialism and subsequently reflexively understanding how the influence of external colonial forces have shaped their lived experience was a tactic through which people identify and acknowledge the existence of their own colonized identities [566]. Through this tactic, people are able to come to terms with their colonized histories and work towards rediscovering and recovering agency over their own hybrid identities.

### **5.3.1.2 Resistance against Collective Colonial Trauma: Drawing on Local Perspectives to Recover from Trans-Generational Psychological Domination**

The users of BnQuora engage in discussions about their history to understand the sociopolitical and economic impacts of British colonial rule over the Indian subcontinent by presenting facts and figures. Specifically, people drew on regional perspectives to challenge their internalized colonial trauma [198, 199]. Through this tactic, people were beginning to resist the long-term hold colonialism had over their collective consciousness as a way in which they were recovering from colonial trauma.

Users of BnQuora, suffering from colonial “traumatic tradition” [199], often collectively exhibited colonial mentality in their conversations. In many cases, we see conversations reflect a kind of trans-generational psychological domination—this view of ethnic or cultural inferiority has become deeply internalized—it is part of people’s identities:

*Although there were some dark aspects of British rule if the British had not come to the subcontinent, India would have been a less developed and backward country than Africa. ... It seems to me that if the British still ruled the Indian subcontinent, India would have been the best country in the world, not America.*

This quote illustrates the internalization of colonial mentality and how it mediates people’s thoughts and perspectives. Here, this user’s views are deeply connected to long-standing colonized

history, whereby they are both shaped and reaffirming a white savior ideological narrative. Whereas this comment illustrates how people have formed a psychological alliance with their previous colonizers, other BnQuora users collectively work to resist and help others resist and subsequently heal and recover from colonial trauma:

*It is said that modern British people forget about their colonial past, but today, I am realizing how true the same is for the people of India ... after reading the answers of most Indians and even Bangladeshis to the questions about British rule. Most of the answers I see are typical examples of Stockholm syndrome. Today, the citizens of the Indian subcontinent are praising the British for giving us democracy, railroads, and the rule of law, who once plundered the whole of India for hundreds of years.*

This comment illustrates how colonial mentality in the Indian subcontinent is the equivalent of Stockholm Syndrome—a condition where the victims of abuse or kidnappings develop an emotional attachment to, and psychological alliance for, their oppressors [159]. Other postcolonial studies have used this concept to describe the attitude of the previously colonized populations towards their colonial rulers [183]. Similarly, through this user’s observations of the interactions of many users of BnQuora, people’s identities as being mediated by long-standing and internalized colonial mentality. While some users on BnQuora engaged in conversations highlighting an acceptance and celebration of British colonialism—thus exhibiting colonial mentality—other users were actively working to help them see the dark realities of British colonialism as a collaborative effort to help the former group recover from colonial trauma.

In this view, some users are challenging narratives that continue to perpetuate and strengthen colonial mentality amongst other members of the forum. As historic trans-generational trauma is deeply embedded in the discourses on BnQuora, many users are aware that this narrative was constructed by those with power—in this case, the colonial rulers—and these narratives have continued to dominate and marginalize people in the Indian subcontinent:

*In this way, the British showed to the outside world as though the rule of India were a huge burden for the British, they were just running the country of the Indians out of their sense of grace despite their own detriment.*

Here, the user is describing the view that colonizers perpetuated about their colonized subjects. The myriad discussions amongst members of BnQuora highlight how colonial influence on people's identities and how they continue to perform and see themselves have persisted and systematically marginalized the voices of the Bengali populations of the region. Through their conversations on BnQuora, users work to understand the underlying cause of colonial mentality. For example, as educational systems were reformed during British colonialism, colonial ideology has become systematically integrated and deeply normalized, as illustrated by the following comment on BnQuora:

*Q: Why did the British never teach their sanguinary colonial history and the horrible things they did in India for 190 years in their schools?*

*P1: What you think is terrible and gruesome is not what the British think. Otherwise, even after so many years of independence, they talk proudly [about their act of colonialism].*

Beyond understanding colonial mentality, we also find that users of BnQuora are tactically drawing on local perspectives and histories to challenge colonial mentality. They take it upon themselves to help others recover from this colonial trauma. Whereas colonial mentality makes people feel inferior or incapable of doing anything on their own, we also see that others were trying to illustrate the reverse—that the colonizers were seeking them out because of the strength of their own regional resources and practices. We see users of BnQuora employ this tactic across many threads. For example:

*Q: Many people say that while the British took away a lot of resources from India, they also gave a lot. - How valid is this claim?*

*P1: [You] can judge for yourself whether the British took more resources from India or not. ... The muslin that survived the rule of all the empires in Bengal for thousands of years came to an end just a few decades after the arrival of the British. ... The salt business of the whole of India was turned into a monopoly by the British. ... I don't know if you would believe it or not - from 1765 to 1937, Britain is estimated to have embezzled 45 trillion pounds (17 times the current GDP of Britain) from the subcontinent.*

Here, by presenting historical accounts from the British colonial period in the subcontinent from the point of view of the colonized local and indigenous people, the user is reclaiming agency over determining how they want to be viewed in the global narrative about colonialism. In sum, the users of BnQuora engage in a collective reflexive conceptualization of colonization in the context of the subcontinent. Here, we underscore how people are engaging in tactics to both understand and help others recover from colonial domination over their minds and identities.

### **5.3.2 Mourning to Continue Healing from Collective Trauma: Speculative Identity Play**

This phase of mourning often overlaps with and can accelerate the previous phase of re-discovery and recovery. We saw in the previous phase, people were actively resisting colonial mentality—rediscovering their concept of self from the traumatic mark of colonialism. However, as colonization has made many irreversible changes to the global socio-politico-economic structure and distribution of resources [17, 198], for some aspects of identity, perhaps there might not seem to be an alternative to the present condition. In such scenarios, our findings reveal that people engage in mourning [330]. Mourning is an important step towards healing. On BnQuora, we find that the users collaboratively engage in speculative play as a tactic for mourning and subsequently processing their self identity and thus reclaiming narrative agency.

To make sense of their changing identities, we find that many users of BnQuora are engaging in what we dub as speculative play. Whereas identity play is the process through which people assume different roles or perspectives to make sense of their own identities (e.g., playing as a Democrat and a Republican to make sense of one's own political identity), we find that people are engaged in a kind of playful speculation where they theorize about their colonized histories and identities, mourning what their lives could have been had colonization never happened.

For users of BnQuora, speculative play became an important mechanism for processing their colonized identities. Here, they engage in speculating about how their lives would have been different had they not been colonized, or if colonizing rulers had never left their region:

*Q1: If the Indian subcontinent were still part of the British Empire, what would it be like now?*

*Q2: What would India have looked like if the British had not ruled India?*

Here, the users are engaging in speculative identity play with regard to crucial political events of colonial history. At this point in time, the undoing of the arrival of colonial power and the subsequent colonial rule over the subcontinent is not possible. The damage of local and indigenous cultures and the looting of resources done during the colonial period are unlikely to be reversed. Speculation, as a kind of mourning, supports them in analyzing different points of their history, effectively accelerating the previous phase of rediscovery.

Being able to ask questions through which people could playfully and collectively come together to engage in collective speculation was a vital tactic in service of reclaiming narrative agency over their own identities. If we take the example above, this initial question prompted a lengthy discussion whereby people came together to collectively deconstruct their colonized histories and identities. While many of the conversational threads discussed the myriad contributions British colonialism made to the social systems mediating the routine experiences of the peoples of the Indian subcontinent (e.g., educational systems, railways, and telegraph systems), these questions also afforded people an opportunity to discuss the overwhelmingly “gruesome” impact of colonization

on their social worlds and opportunities. For example, we see discussions about how colonialism worked to create disadvantages and limit the advancement of the subcontinent. People came together to discuss how colonialism actually obstructed opportunities for growth within the region. As illustrated by the following example:

*If they [the British colonial rulers] never came, the Industrial Revolution would have taken place here, not in Britain. ... This would have been one of the wealthiest and the most prosperous countries in the world, if not the wealthiest. It could have been a superpower through industrialization, economic and military forces, ... rich in history and traditions, the breeding place of science and education.*

Through speculative play, we see that people are collaboratively mourning and thus processing many aspects of their colonized identities. Even years after colonial rulers left the subcontinent, the impacts of their policies and practices remain deeply embedded in people's lives. The various dimensions of this marginalization include political subjugation, economic extortion, uprooting social structures, and inducing an inferiority complex among the native population [199, 311, 545]. Through their speculative questions and ensuing conversations, people are able to come together with like minded others to mourn the collective trauma of colonialism and work towards reclaiming agency over how they see themselves and want others to see them.

### **5.3.3 Dreaming Towards Unity: Reconfiguring Geopolitical Relations**

Fanon argued that after independence, a reformation of economic, social, and political structure is required to foster a truly independent national community [198]. To move in that direction, possibilities are to be expressed, debated, and constructed. These dreams need to be based on the local and indigenous people's exploration of their own culture and social order. Dreaming is considered the most crucial for decolonization and must be allowed to run its full course, meaning that people should be allowed to engage in dreaming through candid expression, debate, and consultation without any external constraints or influence [330]. Given how colonialism had shaped the

region— through the partition of Bengal, the annexation of East Bengal (then named East Pakistan) to West Pakistan on the sole basis of religion, and subsequent Pakistani colonial subjugation until gaining independence as Bangladesh—dreaming would include re-evaluating the social, political, economic and judicial structures of the region, as well as reflecting the values and aspirations of the Bengali people.

Unlike many other online spaces, Quora encourages and supports discussions on political issues in addition to a range of other topics. While being demographically diverse across different nationalities, religions, and gender identities, prior work has found that the users on BnQuora often share a strong sense of unity based on their ethnolinguistic background [149]. We find that users of BnQuora, in particular, are engaged in conversations that scrutinize and seek to reconfigure geopolitical divisions—the political divisions that have come to divide them along geographic dimensions. Through the opportunities to exchange diverse perspectives, users of the platform are dreaming and re-imagining their societies such that they move from being divided to being united as a tactic through which they are working to reclaim narrative agency.

Due to the harsh geopolitical divisions that had come to characterize their colonized identities, people from India and Pakistan would not normally engage in conversations with people from the other nation-state, especially not in productive or constructive conversations [223, 393]. Yet, for users of BnQuora, as they started collectively reconstructing their identities, this also meant that people were working with other people, irrespective of national, geographic, and religious identities, to imagine a better future for the region. Users engaged in discourses about the broader implications of various colonial historic events and their effects on contemporary political issues. Through participating with users of different nationalities and perspectives in an aggregated Q&A thread, discussion served as opportunities through which people could deliberate how the impacts of colonial rule were mediating their modern-day geopolitical discourse. For example, users from different countries came together to collectively explore their views about the colonially incited and long-contested Kashmir issue—this issue is concerned with the withdrawal of Article 370 of

the Indian Constitution<sup>5</sup>. This is illustrated through the following questions posed by a user of BnQuora:

*Q1: How do you see the new decision of the Indian government on Kashmir?*

*Q2: Why does India not want to sit in talks with Pakistan on Kashmir?*

In analyzing the ensuing discourse, we see that other users were coming together to develop a collaborative, unified understanding of the issue, as opposed to maintaining separate ideological stances. These questions served as a jumping point through which people subsequently engaged in conversations about their colonial histories, especially by critiquing the many policies imposed on them by British colonialists, such as the religion-based division that separated India and Pakistan. Users are working to establish a sense of commonality:

*India-Pakistan relations are not good. But this does not mean that as an Indian I have to hate all Pakistanis. It is a very toxic thought and a political tool. Using this tool, political leaders in our countries divert our attention from key issues. As educated persons, we need to understand these twisted tricks. ... You will see a lot of Indian followers of many Pakistani writers on the Quora platform. For example, the names of Mohammad Amir Khokar and Noman Ashraf first come to mind.*

The above quote is an illustration of a user who is actively working to re-imagine the harmful political structures that have emerged as a colonial product. By reflecting on the present reality of the political relations in the region, this user is dreaming about better geopolitical relations. This user is explaining how there exist people who categorically identify as Indian who actively embrace Pakistani writers, highlighting how the divisions between these groups are socially constructed through colonial manipulation. In response to this post, another user from India expressed how

---

<sup>5</sup> Article 370 of the Indian constitution (1954) gave special status to Jammu and Kashmir (a region located in the northern part of the Indian subcontinent) because of the larger region of Kashmir being the subject of dispute among India, Pakistan and China since the partition of 1947 [291]. Its “inoperative” status since 2019 revoked the special status of the region [409].

they were “the same as people from Bangladesh and Pakistan”, while a similar sentiment was reciprocated by users from Bangladesh. Through these conversations, we see that people are actively underscoring the “we-ness” or “one-ness” that is a most evident characteristic of their broader local and indigenous ethnolinguistic identity. They are underscoring how they are all the same as opposed to different.

We also more explicitly observe this shift towards a more connected, unified identity on a broad level on BnQuora. For example, users began to engage in conversations centered around how they might re-imagine their identities in looking toward a more unified future:

*Q1: India-Pakistan-Bangladesh - if these three countries become one again, will it be possible to be the leader of the world?*

*Q2: Is anyone thinking of the unification of the two Bengals?*

However, in general, we found that the users prioritize their national identities while participating in such discussions. They value their countries’ sovereignty over the unification of geographical borders but deeply support the improvement of relationships across geopolitical boundaries.

*P1: We [Bangladeshis] have the opportunity to rule ourselves for the first time in thousands of years of history. No one will want to lose it.*

*P2: I pray to the Almighty that may this bond of these Bengali souls remain and strengthen day by day.*

As we can see here, these users are explicitly dreaming of a reformation towards more collaborative geopolitical relationships in the region. Their participation in this space is being mediated by a reformed, united, decolonized identity and framing their questions and conversations through that revised identity lens as they explore and negotiate various possibilities as a tactic for reclaiming narrative agency.

### 5.3.4 Commitment to Commonality through Anti-Categorical Identity Framing

After exploring different possibilities, in order to continue to work towards reclaiming narrative agency, the previously colonized people have to commit to a single direction to which they want to take their community. However, there is no single way to express people’s commitment to that goal. In this attempt, people need to cooperate with each other and move beyond their communal identities and family histories – their divided identities across various dimensions [330].

Through their use of BnQuora, we find that users are working to establish commonality by seeing themselves through fluid, multi-sectional identities. This is what McCall dubbed an anti-categorical identity framing [372]. People were questioning colonial influence on their identities, and thus, highlighting how those inter/intra-categorical identity framings came to be. The goal of this strategy was to move from thinking about divisive categories and their differences to creating a unified version of themselves. This approach aligns with Mignolo’s argument that “pluriversality of each local history and narratives of decolonization across different intersectionalities can connect through their commonality” [383] which can be the basis of their commitment to the process of decolonization. Finding this commonality does not mean subsuming the intersectionalities and conceiving only one way of decolonization; rather, this process is about unsettling single authoritativeness and seeking cooperation [387].

The decolonization practices on BnQuora served to highlight how colonialism was deeply embedded in how they see themselves and others as being different or the same — the perpetual divisive embedding of categorical identities (e.g., Bangladeshi Muslims, Bangladeshi Hindus, Indian Hindus, and Indian Muslims). On BnQuora, many of the conversations seek to make sense of these divisions and categorical identities. An overwhelmingly large number of discussions focus on the current intercategorical complexity of people’s colonized identities. The following examples illustrate this:

*Q1: Are you first a Muslim or a Bengali or a Bangladeshi?*

*Q2: Are you first a Hindu or a Bengali or an Indian?*

*Q3: Which consciousness works more between Bangladeshi and Indian Bengalis?  
Bengali nationalism, or Indian or Bangladeshi nationalism?*

These and other examples highlight how users are trying to make sense of their differences. We see that people are grappling with the intercategory complexity of their intersectional identities, and through their conversations, it is clear that many users of Quora, as illustrated above, frame questions whereby they inherently privilege certain identities over others.

Whereas many users are working to make sense of the complexity of their multiple and often overlapping identities, other users, in answering these questions, are pushing for a more anti-categorical perspective. We see people draw attention to how identity categories are flawed and incomplete [372]. They explain the intricate, entangled relationship between various identities, such as nationality, religion, and language. For example, a Bangladeshi user shares their personal views about intersectional identity:

*One identity cannot tarnish another. Just like I eat panta-hilsa [traditional Bengali food], I wear punjabi-pajamas and spontaneously celebrate the pahela Baishakh [Bengali new year], and then I fast for thirty days in the month of Ramadan according to the instructions of Allah. I am a proud Bengali and a selfless and contented Muslim. I hold in my heart the culture of this country, the way of life, ... the inspiration of the great liberation war of Bangladesh; I am a Bangladeshi by all means. ... When Sourav Ganguly [an Indian Bengali cricketer] became president of the Indian Cricket Board, I became happy because I am a Bengali. When India plays against Australia in the Cricket World Cup, I wholeheartedly support India because I am also from the Indian subcontinent.*

Through such conversations, users often suggest that the differences in certain identity aspects (e.g., linguistic or religious) should not undermine other identity dimensions (e.g., national, geographic, or cultural). They are working to diminish the divisive perception of the Bengali identity

created by categorical identities and instead work towards broader identity assimilation. Specifically, they draw on local and indigenous regional cultural values to mediate these conversations as they navigate through a spectrum of intercategorical identities to argue for an anti-categorical identity framing. The users on BnQuora in the dreaming phase expressed their aspiration for unity. To work towards that goal, in this phase, users start to think themselves through frames beyond the boundaries of their communal and national identities and adopt an anti-categorical lens to find commonality among themselves to express their commitment to the objective.

### **5.3.5 Action towards Repairing Sociocultural Structures: BnQuora as an Identity Mirror**

The fifth phase in the process of decolonization is action [330]. It constitutes concrete steps toward the dreamed objective to which all in the colonized community had committed themselves. The users on BnQuora use the platform as an opportunity to repair their sociocultural identity by communicating and learning about each other.

After British rule and what emerged in different phases spanning 1947-1971, the Bengali people were separated by the borders of Bangladesh, India, and Pakistan [78]. This prolonged separation has created a range of communication gaps, unawareness, stereotyping, and misunderstandings. This isolation has been especially hard on Bengali culture, as the fissures caused by colonialism have created a cultural vacuum. In decolonizing their identities, we see users on BnQuora revisit their regional cultures in an attempt to repair the sociocultural structures that have deteriorated. Traditional mainstream media has shortcomings in bridging that gap. What was once a community with a common social and cultural background living within the same borders, geographic migration served to disconnect these peoples. BnQuora users have complained that mass media organizations in Bangladesh often do not mention the substantial cultural similarities between Bangladesh and north-eastern Indian states or the sizable Bengali population in Pakistan. Through BnQuora, users were able to learn more about the lives of Bengali people in other countries, as we can see in the following conversation:

*Q: I have heard that there are many Bengalis in Karachi [a large city in Pakistan].*

*How is the city of Karachi?*

*P1: The city of Karachi is very beautiful. Many Bengalis live in this city. Due to the use of the Bengali language in the city, our [the Bengali people's] movement is not a problem.*

Similarly, across the various threads we analyzed, we found several emergent discourses where Bangladeshi users were being exposed to the experiences of people outside of this region for the very first time. They were interacting and communing with Indian Bengalis from Assam and Tripura, to name a few. Whereas Bengalis were learning about the experiences of others, the BnQuora platform also enabled them to learn about the variations of Bengali culture and language in other countries. Given how religion was the primary mechanism through which the three countries came into being, religion became a topic of interest that people discussed.

As the Bengali people from different countries, religions, and nationalities got to know each other, the Bengali culture, which was often marginalized during the colonial rule in the Indian subcontinent, is made visible through the emerging sociotechnical structures of BnQuora. The native language is being normalized and integrated as part of the shared Bengali ethnolinguistic identity of BnQuora. We see this in how users address each other as *bhai/dada* (meaning brother) and *apa/didi* (meaning sister). Some users described how, through BnQuora, they were finding their “long-lost brothers who found each other after a long time”, given their long separation, getting connected on this platform, and the role of Bengali culture in their unification. In this way, people are using Quora as a medium for exposure to Bengali culture, which contributes to the decolonization process. A user expressed this by saying:

*However, if I keep aside the rules and politics of Bangladesh and West Bengal [an Indian state], then be it emotion, eating habit, culture, manner, practice, I think both peoples [from Bangladesh and India] to be the same. I pray to the people to strengthen the positions of the two Bengals through the exchange of culture.*

On a broader level, users across BnQuora who have been disconnected from Bengali culture are seeing themselves reflected in the people from Bangladesh—the platform is acting as an identity mirror. As explicated by an Indian user, who described the stark similarities between Indian and Bangali identities:

*But hopefully, these ideas are slowly changing. For the benefit of some influential people and organizations, today we are watching [some Bangladeshi media celebrities], and maybe thinking to ourselves, they are Bengalis just like us.*

As BnQuora users engage in these discussions, many of them call for social, political, and economic reformation, which is an important aspect of the decolonization process [198]. According to Laenui, articulating these plans through media can be an effective action towards decolonization [330]. Based on the community’s dream for improved geopolitical relations and a commitment to highlighting commonality, the following user shares their idea of action on BnQuora, saying:

*We need to increase cultural exchanges and mutual economic transactions, as well as make the passport visa issue more accessible.*

In sum, through their interactions on BnQuora, users are more broadly adopting their decolonized identities as a community, which reflects their identity, values, and bonds as members of the Bengali ethnolinguistic group. This inspires BnQuora users to call for repairing what is a fractured and fragmented sociocultural structure, and their call for action is mediated by opportunities on BnQuora for reclaiming narrative agency.

## 5.4 Discussion

In this chapter, we have explored how the Bengali people, an ethnolinguistic group with a history of suffering from prolonged colonialism, are actively decolonizing their identities on BnQuora

and thus actively working to reclaim narrative agency over their own identities and histories. Drawing on Poka Laenui’s framework of decolonization [330], we have found that the users of BnQuora grapple with the concepts of colonialism and its impacts. In discussing their lived experiences and engaging in speculations, users of the platform lament the impacts of colonialism over their narrative agency and assist each other in making sense of and articulating the social, economic, and political harms that colonialism has created. Through this process, they resist transgenerational colonial trauma and reclaim ownership over their own identities [199]. Finding commonality among themselves through their broader Bengali identity, people from different intersectional backgrounds explore the possible directions toward decolonization. Using BnQuora as a media, they call for action to reform sociocultural, economic, and political structures—in Fanon’s view to decolonize [198], operating under the landscape of nation states Bangladesh, India, and Pakistan.

A major finding of this research is that people were able to come together and reclaim ownership over their own identities—or narrative agency. One of the primary mechanisms through which people could reclaim ownership over their identities and untangle themselves from colonial narratives [383] was through narrative construction. Our findings illustrate how users challenge the dominant narratives about British colonialism’s “civilizing” mission, helping people understand how colonial structures and norms were taking away control over having agency over their own identities. By engaging in these practices, people were working to subvert the colonial mentality that had continued to mediate people’s everyday lived experience, in continuing to see themselves as subhuman and unable to think and act for themselves. Through their conversations, they were reclaiming ownership of their identities—or generating narrative agency—in identifying the ways in which colonizers undermined local and Indigenous knowledge, fractured communities, and robbed the region of resources, leading to the stagnation of the Bengali people.

HCI and CSCW scholars have described and illustrated the efficacy of expressive writing practices [357]. For example, prior work has found that for people experiencing harassment, writing can be therapeutic [438]. Studies have also highlighted how narrative practices empower users in developing community, support seeking, and countering stigma [13, 187, 338]. Similarly, our work

underscores how collectively generated narratives, as we see in the context of people describing their personal experiences with colonization, can serve as an avenue through which people can work to decolonize their identities. We see this kind of work as a form of narrative resilience. Resilience is defined as how people bounce back from threat or vulnerability [366]. Narrative resilience, then, refers to the ways in which narrative practices serve as a collective and reflexive mechanism through which people work to generate resilience. Much like art therapy [349], which is a form of therapy that encourages victims of trauma to engage with the conceptual issues of trauma recovery through material engagement and expression as manifest through art (i.e., painting and drawing), we see that through collective narrative construction, people are engaging with the pragmatic and conceptual issues of colonization and decolonization. While some [565] criticize social justice activities as a “metaphorization” of decolonization, other prominent decolonial scholars [102, 312] have described such creative narrative construction (e.g., writing, poetry) to be crucial to the process of decolonization. Specifically, narrative recovery serves two primary purposes: (1) reflexivity in exploring identities and (2) resisting and re-configuring existing social systems and arrangements.

First, the narratives produced by users of BnQuora explored and tested new conceptual relationships as a means for building resilience. For example, by playfully speculating about their identities or working to shift from inter-categorical to fluid anti-categorical identities [372], we see that these practices serve as moments of self-reflection through which people can reassess themselves and their own self-concept. Through these narratives, by exploring and testing new conceptual relationships, people were working to reclaim narrative agency over their identities.

Moreover, the narratives produced by users of BnQuora could also be seen as a form of resistance. To better explore this phenomenon, we turn to Snow [530], who describes how being anchored in the “we-ness” of shared attributes (e.g., the Bengali identity) and the contrast with the “others,” in this case, the colonizers, collective action often challenges dominant routines and logics through acts of resistance. Specifically, these acts of resistance often manifest as counter-narratives, where people work to revise dominant ideologies and perspectives instead of remaining docile and

vulnerable to those. Our findings illustrate how BnQuora users, such as through speculation and dreaming, echo HCI scholars' argument about the importance of speculation in the liberation of the oppressed groups [52, 345]. Through their exploration and speculations of possibilities about sociopolitical reformation in the dreaming phase, BnQuora users propose and imagine a future where colonial influences are deconstructed and invalidated.

Taken together, the narratives people produce are simultaneously building resilience by enabling people to reflexively decolonize their identities while also dismantling colonialism more broadly and working to revise and renegotiate the separatist ideologies and present-day sociopolitical structures that have shaped the region. Both practices were dualistically contributing to the processes through which people were working to reclaim narrative agency.

## Chapter 6

### Governance Shaping Platform Identity

In building on scholars' perspectives who have started to pay attention to the ways in which broader sociotechnical systems are shaped by and through coloniality [17, 175, 180], we seek to address the broader question of how sociotechnical systems can impede conversations about colonized histories?

In this chapter, we use trace ethnography [222, 423] as an inductive approach to explore people's experiences using Bengali Quora (BnQuora) to learn about, understand, and engage in Q&A-based conversations about colonized histories. We find that sociotechnical systems' design can embody coloniality that mediates people's experiences in adverse and problematic ways. Specifically, our findings underscore how the broader colonial identity shaping the Global South gets reflected in the norms that mediate people's interactions with and on BnQuora and highlight the myriad ways in which this sociotechnical coloniality marginalizes groups of users based on various identity characteristics like religion and language. To do so, we develop a conceptual frame that draws on concepts of performative identity [100], governance [232], and surveillance [212]. Specifically, our findings reveal that Quora's governance structures and content moderation practices give rise to platform identities—platforms can come to privilege certain identities while pushing other identities to the margins, thus reinforcing hegemonic values and norms.

## 6.1 Background

This section discusses how people express and perform identity, especially through languages, and how certain identities emerge as normative while some others are marginalized. We also explain how governance through surveillance reanimates similar dominance of certain identities in online platforms.

### 6.1.1 Performative Identity and Linguistic Marginalization

To explore this question, we draw on the social interactionist view of identity, which views identity as an emergent product of interaction [93, 234, 268]. The formative work of Erving Goffman [234] argues that identity is constructed in relation to social context. He develops this view through a dramaturgical lens, arguing that people draw on extant rules and norms within the social settings they are embedded to perform their self-identities. This social interactionist view of identity exemplifies the strong relationship between individuals and the societal context as a whole.

Moreover, Goffman's work [234] asserts that people performing in relation to these extant rules and norms is the correct way to act and that acting outside of this constitutes deviant, inappropriate behavior. In simultaneously building on and critiquing Goffman's work, Judith Butler [99, 100] draws on gender performances to illustrate how societal power is generated by and through the construction of normative identities and misperception of identities (e.g., gender, nationality, sexuality, and race). In this view, societal constructions determine which identities and identity performances are considered normative and non-normative. Drawing on a gendered performance lens, she argues how gender is not a fixed identity. Rather, gender is constructed and learned through everyday performances (e.g., speech acts and non-verbal actions) in different social settings and is thus modeled after normative cultural and societal logics. For example, in some cultural contexts, women learn that certain languages and styles of dress, such as wearing skirts and makeup, are feminine. This work underscores how identities are political and used as a way of creating power differentials in society.

On a broad level, the work of Butler highlights how members of various collective identities, such as racial, ethnic, and religious identities, learn to perform and maintain those identities as mediated by normative social, cultural, and institutional logics. People from myriad racial and ethnic identities perform the specific language (e.g., regional dialects and marriage ceremonies) and symbolic rituals that continuously re-construct and maintain those identities. It is through these performances that individuals become perceived members or maintain membership in racial and ethnic groups—they come to learn the history, language or dialect, and symbolic systems like religion and cuisine that give them membership into various identity categories. Yet, despite the logics that often mediate how people enact, perform, and maintain their identities, these performances can conflict with how individuals see themselves—their self-concept. The social settings can even threaten the internalized version of themselves, making it difficult to enact and perform that identity [234]. In other words, people can be marginalized or excluded to the periphery of society and denied their place in it based on myriad dimensions of their self-identities, such as their gender, ethnicity, race, nationality, and more [175, 192, 491]. For example, for Lesbian, Gay, Bisexual, Transgender, and Queer (LGBTQ+), the experience of coming out—the disclosure of one’s sexual and/or gender identity—can be stressful and even traumatic, as that identity faces verbal and physical discrimination across social settings, such as home and school environments, and mass media representations [140, 187]. When inflicted through language, this discrimination often takes the form of bullying, hate speech, and threats.

Taken together, language is a central feature of people’s identities and marginalization. In our work, we define language as the expression of culture: a medium through which we perceive and understand reality [579]. Language can create vulnerability as the use of language can exercise power over people by “othering” them [98]. In this view, language is used as a way in which to create power differentials in society, where it establishes who is considered normative and non-normative. For example, in societies with a national language, such as English in the United States, those whose primary language is not English may feel alienated and like non-normative members of society. During British colonialism, various non-Indic words became normative within

everyday conversational vocabularies in the region over time. The use of language becomes a way in which colonizing forces exercise power over their subjects (e.g., English’s dominance in the Indian subcontinent).

While language expression and use can take various forms, such as speech, text, symbols, and hand gestures, this study explores how language shapes the norms of online community spaces.

### **6.1.2 Governance as Surveillance: Towards Platform Identity**

The uses of language in online spaces are often mediated by and through sociotechnical mechanisms of governance—the structures that mediate interactions in digital spaces [522]. Governance can shape the underlying norms of a community in ways that enable or threaten people’s identities and identity expression. Here, we focus on one explicit kind of governance: content moderation. Content moderation is a mechanism of governance through which people’s performances in digital spaces (e.g., the content being produced by people in digital spaces) are established as acceptable or not [232, 464, 469]. The Internet gradually gave rise to a combination of human and machine (e.g., algorithmic) moderation processes through which rules and norms were created and enforced [231, 232].

On a broad level, moderators are often considered the custodians of the Internet—they scrub the Internet of harmful content [232, 464]. In this view, moderators play an instrumental role in creating and sustaining supportive community spaces conducive to fruitful engagement while also playing the role of “protector” in eliminating dangerous content like hate speech and fake news. In this way, moderators can also come to hold a lot of authority and power in the spaces they govern. This is best illustrated by Seering and colleagues [495], who used metaphors like “dictator,” “governor,” and “judge” to describe moderators.

In examining online communities’ governance, CSCW and HCI scholars have explored digital spaces mediated by transparent governance systems, such as Reddit [172, 295]. Early work on moderation, in drawing heavily on qualitative methods, discussed the power structures manifest in large-scale moderation practices on online platforms like Usenet, multi-user dungeons, internet-relay

chat, mailing lists, and more [452, 525]. Other scholars have focused on user-governed communities like Wikipedia [33, 436], Slashdot [334, 442], and free open source software communities [463, 584]. These studies have drawn on quantitative methods to explore the ways in which online behavior is shaped and regulated by and through moderation and how moderation contributes to the success and growth of online communities [333, 520]. Moreover, scholars have examined the evolution of moderation practices, highlighting the different philosophies of moderation over time [119, 176, 213, 321]. Recent works on moderation have explored how different kinds of moderation impact the platform's organizational, legal, and technical standpoints [232], how moderation practices and user-behavior mutually influence and inform one another [207, 224, 403, 496], and how moderation impacts the very people who engage in this often volunteer work [174, 175, 304]. Yet, not all systems govern in transparent ways—prompting the question of how might a lack of transparent governance shape the underlying characteristics of a sociotechnical system and, in turn, enable or threaten decolonization processes and associated identity expression?

To explore this question, we draw on Michel Foucault [212] who, in *Discipline and Punishment*, describes how structures of governance shape social and societal norms in modern societies. To formulate his argument, Foucault draws on the concept of the panopticon, as developed by Jeremy Bentham [59]. The panopticon refers to a specific prison design centered around a tower where the prison guards can observe all of the surrounding cells and, in turn, the behavior of the inmates who are imprisoned. Yet, the tower is designed to keep the inmates uncertain about whether or not they are being observed. Thus, the presence of surveillance mediates the behavior of inmates. Not knowing they are being watched, prisoners engage in obedience through self-regulation and thus discipline themselves.

The perception of being under constant surveillance can become deeply embedded into the social fabric. Certain behaviors and performances are normalized while others are rendered non-normative and marginal—what we refer to as panoptic performativity. Beyond prisons, for example, in a social setting, what behaviors are deemed professional or modern is dictated by the people in the position of power (e.g., colonially imposed ways of dressing over the ethnic attires). However, these

standards are actualized not through any repressive power but rather through normalization and surveillance so that people undertake the ritualized performativity on their own. The individuals who do not conform to those rituals are denied access to places, such as an Indian getting her entry rejected to a local golf club for wearing traditional attire [165].

In the context of digital spaces, the mechanisms of governance can also reflect this panoptic performativity whereby they serve to normalize certain performances of identity and threaten other performances of identity. Importantly, governance extends beyond the role of moderators. That is, there are various social and technical mechanisms that shape, regulate, and thus normalize online community behaviors. For example, through the various sociotechnical mechanisms of Reddit, such as the karma system, upvoting and downvoting, and moderation practices, the site supports and reflects the values of the majority of the users and serves to marginalize the performances of other users of the platform [229]. For example, Gilbert [229] found that “the default masculine whiteness of Reddit” can marginalize women among others. In this way, governance, or surveillance, is mediated by various sociotechnical mechanisms that comprise online community spaces.

In this chapter, we are interested in Quora’s governance systems. It has been speculated that the platform has an anonymous body of human and machine moderators that govern the site<sup>1</sup> . <sup>2</sup> . Here, we focus our analytic attention on the conversations emerging through one of the supported languages as part of this platform—BnQuora<sup>3</sup>—which was designed to enable conversations amongst people using the Bengali language. Through a trace ethnographic analysis [222, 423] of Q&A threads on BnQuora, findings reveal that the governance structures of Quora give rise to a platform identity—that is, platforms can come to privilege certain identities while pushing other identities to the margins, thus serving to reinforce hegemonic values and norms.

---

<sup>1</sup> <https://qr.ae/pGpbFT>

<sup>2</sup> <https://www.quora.com/>

<sup>3</sup> <https://bn.quora.com/>

## 6.2 Methodological Details

In this section, we will provide details about our data collection (e.g., preparing the list of search phrases/keywords) and data analysis approaches.

### 6.2.1 Data Collection

To collect data that could help us understand the ways in which systems of governance mediate people’s identity work and identity expression across national and religious identities in the context of the Global South—a region that was shaped by colonialism—our data collection process combined purposive sampling [555] and snowball sampling [236]. We collected data from BnQuora from 15 May 2020 to 15 July 2020.

In using purposive sampling [555], we created an initial list of terms that focused on identifying features of the platform (e.g. moderation and stages), narratives describing how people were experiencing governance, and different terms focused on the potential identities of the users across various dimensions, such as linguistic, national, or religious. While preparing this list of search phrases, we remained aware of and used both the archaic and revised spellings of our list of keywords (e.g., বাঙালি and বাঙালী for the same word “Bengali people”), and widely used synonyms and similar phrases (e.g., ভারত, ভারতবর্ষ, and ইন্ডিয়া all mean “India,” where the first two words are endonyms for the country whereas the last is the exonym for the country).

Table 6.1 lists the keywords and phrases we used to generate our purposive dataset. Though the quoras API [151] supports keyword/phrase searching directly, it discourages the use of this function. To abide by the advisory *robots.txt* file of the BnQuora platform, we searched for the Q&A threads on the BnQuora platform containing our list of keywords or phrases directly using a web browser. Using this sampling approach, we collected 178 unique Q&A thread URLs. We then passed the links of the returned threads to the API to retrieve and store details about those Q&A threads. Abiding by Quora’s terms of service<sup>4</sup>, we are not allowed to make our Q&A threads-

---

<sup>4</sup> <https://www.quora.com/about/tos>

dataset public. However, using the keywords listed in Table 6.1, our data set should be easily replicable.

Table 6.1: List of keywords and phrases. The search keywords/phrases are not sorted in any particular order.

Keyword/Phrase	Translation and Explanation (if needed)
Quora; Quora (কোম্পানি); Quora (পণ্য)	Quora; Quora (Company); Quora (Product)
Quora বাংলা	Quora Bengali/ Bengali Quora
Quora সম্প্রদায়	Quora community
Quora নীতিমালা	Quora policies
Quora ব্যবহার করা; Quora বাংলা ব্যবহারকারী; Quora লেখক; Quora ব্যবহারকারীদের মতামত	Using Quora; Quora users; Quora writers; Quora users' opinion
Quora এডমিন; Quora বাংলা মডারেশন; Quora বাংলা মডারেটর	Quora administrators; Bengali Quora modera- tion; Bengali Quora moderators
Quora মঞ্চ	Quora stages
Quora ত্যাগ	Leaving Quora
বাংলা; বাঙালি/বাঙালী	Bengali (language); Bengali (people)
ভাষা; আঞ্চলিক ভাষা; জাতীয়তা; জাতি; ধর্ম	Language; Regional language; Nationality; Na- tion; Religion
পূর্ববঙ্গ; পশ্চিমবঙ্গ; বাঙাল; ঘটি; বাঙাল ও ঘটি	East Bengal (modern day Bangladesh); West Bengal (Indian state); Bangal (a term to address the people of then-East Bengal); Ghoti (a term to address the people of West Bengal); Bangal and Ghoti
হিন্দু; মুসলিম; ইসলাম	Hindu (people)/Hinduism (religion); Muslim; Islam

**Table 6.1 continued:** List of keywords and phrases

Keyword/Phrase	Translation and Explanation (if needed)
বাংলাদেশ; বাংলাদেশি/বাংলাদেশী	Bangladesh; Bangladeshi
ভারত/ভারতবর্ষ/ইন্ডিয়া; ভারতীয়/ইন্ডিয়ান	Bharat (endonym of the country)/India (exonym of the country); Bharatiya (endonym for the people of India)/Indian (exonym for the people of India)
ভারতীয় উপমহাদেশ	Indian subcontinent

To increase the breadth and volume of the dataset, we used a form of snowball sampling [236] where we included 625 additional related Q&A threads recommended by the Quora Recommendation Algorithm (QRA) [614]. The keywords-based and QRA recommendation-based datasets were merged to create the final dataset of 803 Q&A threads we used for analysis (see our discussions on this approach’s methodological considerations and implications in chapter 10).

### 6.2.2 Data Analysis

We used an approach from grounded theory [547] for our data analysis. This approach is widely used in qualitative research [174, 503, 569]. For the organization of our data analysis process, we used a qualitative data analysis software called Quirkos<sup>5</sup> to code the Bengali Q&A threads collected from BnQuora. As outlined by Strauss and Corbin [547], we engaged in a process of open coding, where we identified the concepts that appeared repeatedly in the data. Some examples of the codes that emerged in this phase are: “anonymity of the moderators,” “use of upvotes and downvotes,” “difference in dialects,” “difference in adoption of foreign words,” “non-transparency of moderation.” We then collaboratively engaged in the process of axial coding where we combined the open codes to create higher conceptual themes. For example, open codes like “difference in dialects”, “difference in adoption of foreign words,” “difference in use of synonyms” were merged

<sup>5</sup> <https://www.quirkos.com/>

to create the axial code “difference in language practice.” Finally, in the selective coding phase, we highlighted the relationships among the axial codes, which led to the themes we present in this chapter. We collaboratively conducted the preliminary coding of the data<sup>6</sup> of multiple exemplar quotations for each of the emergent open codes. We carefully translated quotations from Bengali to English by translating each quotation multiple times and checking across the translations as a mechanism for internal validity [343, 368]. We performed a reflexive grounded theory-based data analysis [233, 547], which does not call for an inter-coder reliability score [374]. Our familiarity with the sociocultural norms of the Bengali community, as well as with the BnQuora community norms, made me aware of the sociomaterial context of this study. It is important to note that while our work stayed true to ethnographic tradition, the context of the work and its focus on conversations about colonial histories is inherently political. Moreover, and as is an important part of the ethnographic tradition, the unique identities and reflexivity of the researchers have shaped the interpretations of the data that are presented in this work.

### 6.3 Results

On the surface, BnQuora serves as a space for the Bengali people to practice and perform their linguistic identities. On a deeper level, the governance structures of BnQuora serves to marginalize Bengali as a linguistic identity. In the sections that follow, we first provide an overview of the governance mechanisms that mediate people’s interactions on BnQuora. We then describe how these mechanisms of governance come together and give rise to a platform identity: certain identity performances are normalized whereas others are marginalized.

#### 6.3.1 Sociotechnical Mechanisms of Governance on BnQuora

In this section, we describe the sociotechnical mechanisms of governance that shape people’s experiences and participation in Q&A conversational threads on BnQuora. These features include:

---

<sup>6</sup> In the results section of the chapter, we include the English translations of the users’ Bengali quotations only. However, we will be happy to make the Bengali quotations and the URLs of the corresponding Bengali Q&A threads available on request.

the Quora Recommendation Algorithm (QRA), Quora Stages, the Be Nice, Be Respectful (BNBR) Policy, Upvoting and Downvoting, and Moderation.

### 6.3.1.1 Quora Recommendation Algorithm (QRA)

Like many other social networking and Q&A sites, Quora uses a recommendation algorithm—herein dubbed QRA—to organize the user experience and the kinds of threads users see on their individualized homepages [614]. In other words, the QRA determines what posts appear on a given users’ homepage – what posts they see first, which posts they can see or cannot see at all. Moreover, the QRA learns to customize and recommend content based on a user’s activity, e.g. search history and other activities on the platform. This is best illustrated by the following quote from a user on Quora:

*You will see everything after creating a profile for the first time, but the more answers you read, upvote and share, the more similar questions you will see.*

### 6.3.1.2 Quora Stages

Whereas Quora has traditionally served as a Q&A threaded forum space, more recently a new feature was introduced which was dubbed “Stages.” Through stages, users can create a “stage” for a group of users who are interested in common topics and perspectives. Initially, the opportunity for joining a stage was only available if a user was invited by the administrators of the stage. At present, all users can create new stages or join existing stages by sending join requests which are sometimes subject to the stage administrators’ approval<sup>7</sup>. A user can also share an existing thread’s link with the users in the stage. In other words, Quora stages are somewhat equivalent to subreddits or Facebook groups. Moreover, instead of posting directly to the BnQuora homepage, users can ask and answer questions within a stage. This feature is best described by this representative quote from a Quora user:

---

<sup>7</sup> <https://qr.ae/pGPXJ>

*[Stage] is a new feature where people have the opportunity to build a community based on their own interests and likes and dislikes. In a stage community, you can discuss topics of interest as well as collect different content related to specific topics of interest. ... Some of these stages are: Bengalis' Baithaki Adda (careless discussions of the Bengalis), Islam and Muslim lifestyle.*

Using this feature users can participate in discussions within a group that is shifted away from the primary conversations on Quora, thus shifting these conversations from being more visible to less visible. Specifically, users will only see the content produced in a stage if they have already joined, or subscribed, to that stage. Each stage has its own administrative body, who are elected from its membership base.

### **6.3.1.3 The Be Nice, Be Respectful (BNBR) Policy**

All posts on BnQuora are required to abide by the platform's "Be Nice, Be Respectful" (BNBR) policy, which is in place to ensure that the users treat each other with civility and respect<sup>8</sup>. Whereas this policy is in place, users point out that this policy is highly subjective, and its interpretation and implementation can vary from user to user, thread to thread, and stage to stage:

*It is understood that there is an application of a weapon called "Be Nice Be Respectful" (BNBR) policy. But which is not Nice or which is not Respectful, that judgment is relative.*

### **6.3.1.4 Upvoting, Downvoting, and Reporting**

Similar to some other digital platforms (e.g., Reddit, YouTube), BnQuora uses a voting-based rating system for the posts on its platform. Users who can view posts can express their positive and negative opinions about the posts by using upvotes and downvotes, respectively. Like Reddit,

<sup>8</sup> <https://help.quora.com/hc/en-us/articles/360000470706-What-is-Quora-s-Be-Nice-Be-Respectful-policy>

Quora deploys a computational formula whereby a post’s visibility is determined by the ratio of upvotes and downvotes to determine the order in which it will be presented to users [228]. These voting mechanisms play an important role in fostering a sense of community in culturally diverse online spaces [419].

### 6.3.1.5 Moderation

BnQuora (like any other Quora forum) has a moderation team in place to govern the content posted to the platform. However, unlike other user-generated community spaces like Reddit, the moderation team of BnQuora is anonymous and invisible. Users expressed their frustration about the lack of transparency in moderation and moderators on the platform:

*I don't know who Quora moderators are, which country they live in, or what their religions are! But I want to say something to them. ... In fact, Quora moderation itself does not know when an answer will be deleted.*

It is not clear to the users who the moderators are or what rules they are using to moderate the questions and answers being produced by the user community. Thus, moderators can see any user and their activities at any time but users cannot see the moderators.

### 6.3.2 Panoptic Performativity and Platform Identity on BnQuora

In online question and answer spaces like Quora, language becomes the primary mechanism through which people perform their identities. Whereas BnQuora was designed to support conversations and connections amongst an ethnolinguistic identity category, the analysis of our data highlights the ways in which the broader systems of governance within the platform come together to marginalize the very identity the platform supports. Specifically, the various sociotechnical mechanisms that govern people’s interactions on the platform give rise to a platform identity—certain identities and identity performances are normalized whereas others are marginalized. We illustrate

this through the themes of: (1) moderators as prison guards, (2) collective surveillance as enforcing a majority identity, (3) algorithmic coloniality, and (4) staging as self-imprisonment.

### 6.3.2.1 Moderators as Prison Guards: Surveillance as Privileging

Moderators play a primary role in shaping the norms, and thus the identity, of an online community. Whereas moderators have been described as custodians of the Internet [232] who cleanse the online spaces they govern of harmful content, on BnQuora, the users perceive that moderators are actively privileging certain identities over others. Users with diverse identities raised similar concerns, as they perceive their respective group identities to be marginalized through the practices of moderation. Here, an Indian user expresses their perception of moderation to be biased towards the Bangladeshi users:

*But as long as such answers continue to be deleted, I will assume that Quora moderation is biased towards Bangladesh and its one special community, which is very sad.*

Specifically, moderation practices are shaped by coloniality in that moderators come to re-enforce existing power structures in the region that privilege the dominant national and religious (in this case, often Indian and Hindu) identities over the others in the region. Though Bengali is the native language of Bangladesh and some Indian states, as previously described there are subtle differences in the ways they speak or write the language. For example, certain words have multiple synonyms in Bengali and people in Bangladesh and India use different synonyms in their respective countries more widely than the other. For example, the word “water” has two translations in Bengali: “jol” (জল) and “pani” (পানি). People in Bangladesh typically use the word “pani” whereas its synonym “jol” is more popularly used in India. Importantly, these two synonyms are often associated with different religious groups where “jol” and “pani” are used widely in Hindu and Muslim households, respectively. In the conversational threads on BnQuora, we observed a user who asked the following question:

*“How can I recognize a bottle of safe drinking water?”.*

In writing this question, this user used the word “pani” to refer to water, which is also associated with Muslim households. After writing the question, the moderators on BnQuora replaced the word with its synonym “jol”. In the thread that followed, the original poster objected to the edit, and this disruption served as an opportunity for reflection whereby the users started to question the moderator’s collective national and religious identities, and the identity of the platform at large. This is best illustrated through the following exchange:

***Q:** In the question, it was written “pani”, it was edited as “jol”. What is the problem in writing “pani” in Bengali Quora?*

***P1:** I think it could be because the controllers of Quora are Indian. Don’t know exactly though. Because in India it is called “jol”. ... It is disrespectful to the questioner.*

Here, the original poster asked for a rationale for why their post was edited, and other users participated in a conversation through which they were trying to understand how their posts were being governed. In this case, the users assumed that Quora moderators (or as P1 calls them “controllers”) are from India and thus prioritized the synonym which is used more widely in India. Thus, according to P1, this editing was a reflection of the preference of the Indian controllers’—which are akin to prison guards as per Foucault’s panopticon—who were normalizing an Indian style of Bengali writing on BnQuora. On a broad level, this example illustrates how users of BnQuora often use writing styles and language choices as a mechanism for determining the national and religious identities of other users.

Beyond linguistic variations among the Bengali people, they also practice different religions. Currently, the number of Bengalis who practice Hinduism and Islam are both substantial. Yet, many of the Muslim users of BnQuora have claimed that the moderation of the Q&A threads has made them feel invisible:

*No one can avoid the fact that Quora moderation does not delay even a minute in hiding the posts from Muslims if they bring up something about other religions. On the Q&A threads that hurt Islam, even if it is commented on, [the comment] is hidden.*

Here, the user is expressing their perceived discrimination on the platform. According to their experience, in a comparative discussion among Islam and other religions, a post from an Islamic point of view is deleted, whereas the threads from the points of views of other religions, when they are hurtful towards Muslims, are not hidden. Rather, the complains from those who identify as Muslim are removed instead. In this way, the platform comes to exhibit an identity that gives preference to certain religions while pushing other religions to the margins.

In the absence of transparency and feedback about the moderation decisions, users continue to reflect on the moderator's collective national and religious identities:

*With the overwhelming majority of Indian moderators in Quora, they seemed to be a little more arrogant or autocrat-like. ... They enjoy attacking believers of different religions.*

Essentially, it becomes a question of representation. Many BnQuora users speculate that the moderators, the intermediaries between the Quora platform authority and its users, are often Indians who privilege the users whose identities reflect their own. Moreover, several users engage in political discourses to learn about other people's views and perspectives. Yet, several Bangladeshi users claimed that their posts were removed for criticizing Indian political views:

*Is Bengali Quora rapidly losing Bangladeshi users due to pro-India censorship?*

Thus, the users who perceive a discrimination against themselves disassociate themselves with the platform by becoming inactive on the platform and do not contribute to the discussions. It reinforces the communal division, distrust, and the lack of cooperation among people in the region

that was initially sewn by colonial practices and policies. The moderators on BnQuora often do not provide transparent justification for their moderation work which in this case had led many users to infer the moderators' identities and generalize those identities to the platform's identity. Taken together, through people's performances on BnQuora, their experiences with moderation illustrate a platform identity that privileges certain performances and marginalizes others.

### 6.3.2.2 Collective Surveillance as Enforcing a Majority Identity

Beyond the ways in which moderators are surveilling and shaping people's performances of identity on BnQuora, our findings also highlight the ways in which the users of BnQuora also work to collectively police, or surveil, people's linguistic performances in a way that gives power to certain identities and marginalizes others. Here, we highlight two primary mechanisms of collective surveillance: (1) voting as erasure and (2) hierarchical monitoring.

**Voting as erasure.** Much like Reddit where upvoting and downvoting can serve to re-enforce the values and perspectives of the majority user population [229], we also find that upvoting and downvoting served as community mechanisms that systemically normalized and privileged certain identities while marginalizing others.

Given that upvoting and downvoting contribute to what posts become more or less visible, users of BnQuora observed first-hand how the community itself was shaping the visibility of posts:

*Like the real world, the majority is dominating here. People who are the majority about an opinion here are downvoting, reporting and upvoting the others' opinions in order to establish their own opinion. Everyone's freedom of speech is being protected just like the real world! The difference in the number of active users is being revealed through downvotes and reports.*

Although upvoting and downvoting often reflected people's personal biases and prejudices, moderators of BnQuora use post visibility to make decisions on which posts to promote or delete altogether:

*I've seen you will create some well wishers on Quora as well as a bunch of enemies. They will vote down your answer just because they do not like you. And Quora moderation will also delete the answer by looking at the number of downvotes without reading the answer.*

Thus, sociotechnical artifacts like upvotes and downvotes act as forms of collective surveillance that serve to reinforce power and hegemony in digital platforms. As expressed by this user, freedom of expression is only being reserved for those whose linguistic performances were aligned with the normative logics of the dominant majority of users. For those whose performances did not align, their identities could be erased from the platform altogether.

Even when posts received the largest number of upvotes, if those posts went against the platform's identity, moderators also removed them:

*... Moderators of BnQuora were hurt by that answer. Though my answer received the highest number of upvotes, they took it down to save their faces, but they did not remove the question. Then, I realized that BnQuora is not Bangladeshi, rather Indian, as if [the platform] has a nationality.*

Here, the user is reflecting upon their experience about a previous post. In response to someone's question of "Why do Bangladeshis who never have come to India and do not understand things correctly, comment randomly?", the user replied that "there were some people in both Bangladesh and India who have such mentality and such Indians also comment about Bangladesh without proper understanding or knowledge." When this post was removed, this signaled to the user that the platform itself had a "nationality", and that user posts that did not conform to that platform's identity were subject to erasure.

**Hierarchical Monitoring.** Whereas we previously described the ways in which users were normalizing the kinds of performances on the platform through their voting behavior, this kind of activity is part of a broader system of collective policing that we conceptualize as hierarchical

monitoring. Each individual who is part of this sociotechnical system plays a role in monitoring and regulating the behaviors of others. This is best illustrated by the following quote:

*I am writing anonymously, but some “detectives” will closely analyze my writing and find out my identity.*

Beyond the voting behavior that can lead to the erasure of certain performances that go against the platform’s identity, as part of hierarchical monitoring users are also regulating the kinds of language used in Q&A threads. Only here, users are collectively working to shape people’s performance through harassment and ridicule. For example, as previously described, there are certain differences in the writing styles and dialects of Bengali languages among different regions of Bangladesh and India. Specifically in Bengal (Bangladesh and West Bengal), being a densely populated country, this variation is more noticeable. This is best articulated by the following quote:

*“In Bangladesh, the dialect varies, as the distance varies from 40-45 km.”*

The users on BnQuora recognize this phenomenon. The Bangladeshi dialects can be easily recognized as “different” by the Indian Bengalis. Here, this user is referring to those differences in writing styles and explaining how their writing style can be a lens through which the anonymous moderators can monitor them based on their “extrapolated” identities. Some users have shared their experiences of writing on BnQuora using the phonetic spellings of their regional dialects and being criticized for that. As expressed by the following user:

*Different regions of Bangladesh have different [dialects]. Many people post some ridiculous answers and comments about the languages/dialects of other regions. It is very annoying to me.*

Users on BnQuora have tried to understand the motivations behind such derogatory or disapproving attitudes towards the Bengali language’s Bangladeshi dialects. Some users have concluded

that this attitude is rooted in the sociohistoric perception of the subcontinent. One such crucial milestone in history was the postcolonial partition of 1947. As Bengal was used as a site of partition, both East and West Bengal faced a huge refugee crisis. The refugees from rural East Bengal/East Pakistan (modern-day Bangladesh) being addressed as “Bangals” as a form of harassment:

*The main reason for this belittling may be the oral language, the rural customs of East Bengal. ... The word [Bangal] may have become a symbol of bitter opposition when the stream of helpless people from East Pakistan, after the partition of the country, who were occupying vacant lands, gardens, etc., in Kolkata and other districts. Everyone from East Pakistan has this label in the post-1947 period.*

Similarly, the users’ views towards their history of colonial subjugation and its impacts also influence how Bangladeshi and Indian users view the concept of adopting foreign words in their writing. We found that the Bangladeshi writing style accepts using widely used foreign words (e.g., Arabic, Persian) in Bengali writing, while the Indian writing standard is more conservative in that regard. An Indian<sup>9</sup> user expressed their denunciation towards this liberal adoption of foreign words by the Bangladeshi Bengalis saying,

*... Now let’s talk about my dislike. ... Bangladesh is the country of Bengali language, the country of sacrificing life for language. I personally do not like the use of many Arabic, Persian and Urdu words in Bengali in that country.*

In that discussion, the Indian users discouraged the adoption of words from foreign languages, viewing those as external influences on the Bengali language. These experiences are deeply rooted in coloniality, where religious identity became correlated with social status in Bengal. Similarly on BnQuora, the use of certain words that have come to be associated with specific national or religious

---

<sup>9</sup> We did not collect the users’ information. We infer this user’s national identity from the title of the thread where they were trying to answer the question: “As a Bengali from Kolkata (an Indian city), what makes you sad about the Bengalis of Bangladesh?”

identities provides linguistic cues for a users' perceived "Bangladeshiness" and "Indianness", or "Hinduness" or "Muslimness". This relates to Homi Bhabha's concept of hybridity, where users are harassing those whose language use is a product of coloniality. Through this hierarchical monitoring of the linguistic choices being made by others, people are collectively working to admonish those whose performances counter the platform's identity.

We also see this hierarchical monitoring taking place through tagging, which is one of the features of Quora. Users can assign different tags (e.g. keywords and phrases) to the Q&A threads, where the tags are community generated and used to identify the primary topics covered in any particular thread. Quora creates a page for each of those topic tags to aggregate all the threads using that tag. Through the use of tagging, members of BnQuora are collectively working to create a platform identity that establishes certain performances as normative and marginalizes others.

One of the primary pages on BnQuora centers around the topic of the "Bengali (people)." As part of the page the users have include an image to best represent the topic, where for this particular topic the chosen image was a map of the Indian state West Bengal. Such practices make the default representation of Bengali people on BnQuora to be synonymous with the Indian Bengalis while marginalizing Bangladeshi Bengalis. In other words, Indian Bengali representation becomes the normalized representation of Bengali people. As described by a user:

*Why is only an image of West Bengal's map as the picture of the 'Bengali' topic in Quora? Don't Bengalis live in Bangladesh?*

### **6.3.2.3 Algorithmic Coloniality**

As previously described, through the process of colonization, the Bengali people were separated geographically by religion as a way in which discord could be created and amplified in the region. As a reflection of the impacts of colonization, one of the primary conversational topics across the threads is religion. Given that many users expressed adopting and subsequently engaging in conversational threads on BnQuora to learn about the experiences and perspectives of those who

share the same ethnolinguistic identity, our analysis identified several threads where users were complaining about how the majority of threads being recommended to them by the QRA were focused on problematic and divisive topics like religion. As explained by the following user:

*Why is there nothing in Bengali Quora except religion and caste? You may be new to Quora! (If I am not wrong). Quora promotes the most viewed/popular and controversial topics in newcomer feeds. It is automated from Quora's system.*

By promoting controversial Q&A threads that focus on religion, the algorithm also comes to exhibit a kind of coloniality—what we dub algorithmic coloniality. In this view, the algorithm is promoting and thus continuing to amplify conversations around features of people's identities that have sewn discord for decades. When these threads are promoted to new users, those threads can seed a new users' initial activities and set the stage for a user's subsequent interactions and experiences with the platform.

Moreover, users described how the QRA learns about people and works to identify them. This relates to Cheney-Lippold's [117] concept of algorithmic identity, wherein he argues that algorithms process data to measure certain features about us, such as our race and religion. Such use of quantification and statistical accounts of the world as a tool was one of the characteristics of colonial rule. For example, Herbert Risley, a colonial government official, attempted to understand the racial origins and castes of the people in Bengal (e.g., Indo-Iranian<sup>10</sup>, Dravidian, and Mongoloid) based on the people's physical attributes (e.g., nasal index and skull width) which are heavily critiqued by modern day scholars [217, 462]. In the context of computing, Dourish and Mainwaring have identified similar commitments to reductionist quantitative measures for understanding, evaluation, and prediction as the "colonial impulse" of ubiquitous computing [180]. Only in the context of BnQuora, users express that the platform's primary identity centers around religion:

---

<sup>10</sup> Because of use of atrocities committed in the name of the Aryan race by the Nazis, academics encourage replacing this term with "Indo-Iranian" [608]

*No matter how successful and enterprising you are in your own [professional] field  
- you have only one identity in Bengali Quora - Hindu or Muslim.*

Thus, the QRA is prioritizing users' religious identities and personalizing their experience on the platform around that one identity. To draw on some historical context, in the British-organized municipal elections during the last decades of the Raj, Muslims could only vote for other Muslims, and Hindus for Hindus [245]. Much like how the colonial rulers crudely generalized the diverse population of the subcontinent, the algorithm exhibits a kind of coloniality in that it shifts people's focus from the myriad topics they can discuss and enforces a platform identity centered around religious discord and highlighting differences amongst members of the community.

#### **6.3.2.4 Staging as Self-Imprisonment: From Visibility to Less Visibility**

Due to how the QRA promotes communally divisive threads to its user population and the ways in which several members were being marginalized, BnQuora recently launched a feature called "stage" as a solution to such conflicts. Through their use of stages, users of BnQuora can engage in conversation with like-minded others in a less surveilled but far less visible space. Yet, the use of stages can also be seen as a way to strengthen the platform's identity. That is, as users with non-normative identities or who wish to engage in non-normative conversations confine themselves within those stages, this creates a kind of platform purity within the primary conversational spaces on BnQuora. As expressed by the following user:

*The announcement of the launch of the stage is a new, very good initiative. It is a new feature where people have the opportunity to build a community based on their own interests and likes and dislikes. In a "stage" community, you can discuss topics of interest and collect different content related to specific topics of interest ... Now, you can find the space of your choice and share knowledge. We hope that we will not have to look at any extremist questions, Islamophobia, etc.*

This user had experienced extremism and Islamophobia on BnQuora earlier. After the introduction of the stage feature, such users approach stages as a less visible space where they would be subject to less monitoring. Therefore, the users who find themselves marginalized can confine themselves within stages while discussing their religious identity and practices. Whereas in the context of marginalized identities, finding safety is important [187], in other cases staging can work to further polarize an online space (e.g. for those engaging in political conversations) or serve to make non-normative identities less visible or even invisible.

Foucault described that the individuals who break laws in society do not cease to exist. Rather, they are removed from society within the arrangement of prisons [212]. Similarly, for those users whose identities are marginalized by BnQuora’s platform identity, moving to stages is a kind of self-imprisonment wherein they can engage in discourses with a homogenous group. They do not cease to exist, but are relegated to an alternative space that will not challenge the platform’s normative identity.

## 6.4 Discussion

BnQuora offers an intriguing place to explore how sociotechnical systems may reproduce colonial structures that shape people’s identity performances. Studies have shown how online platforms such as Quora can host a large volume of discussions on regional issues when they are operated in local languages [151, 567]. From this point of view, the launch of Quora in Bengali seems to be a benevolent attempt on the part of the platform designers for creating a more accessible space for the Bengali speaking users. The online platform Quora that runs in English for an international user group and its Bengali counterpart, BnQuora for the Bengali speaking users alone have similar features and scaffolds. The only difference between Quora and Bengali Quora is the languages of communication on these platforms. The setup and the structure of the platform does not regard the social, political, cultural, and historic background of the Bengali people and their language.

Thus, the platform appears to impose a colonial “notion of universality” [180] across platforms of different languages.

But, BnQuora does not bring together one coherent group. It pools different communities with a shared language and a colonial history anchored in a range of temporal horizons, discourse practices, national and religious identities. The Quora recommendation algorithm (QRA) in combination with the upvote and downvote features appear to amplify some of these distinctions. According to several users the recommendation algorithm on Bengali Quora considers participants either Hindus or Muslims – divided on a religious spectrum, above everything else. These divisions tend to emerge on controversial Q&A threads and subsequently promoted by the algorithm to create greater online traffic and user engagement.

Understanding what happens to these distinctions on the platform becomes important if we hope to comprehend the role of sociotechnical systems in reproducing colonial structures that shape people’s identities. The combination of a performative approach to language use and identity expression with Foucault’s notion of panopticon offers a dynamic perspective.

The participants’ identity performances play out on a platform that promotes both visibility and invisibility. Much like Foucault’s depiction of the panopticon, participants’ posts are exposed and visible to all, while moderators are hidden behind a curtain of anonymity (see Figure 6.1). This creates a sense of surveillance among some participants, and in particular, those who are exposed to anonymous moderation practices that erase or change their posts.

In contrast to Foucault’s prison, poky cells do not define the BnQuora platform. Free to roam, participants can contribute to any discussion they please. When a user does not behave according to platform norms, they usually do not receive any conclusive or tangible punishments like getting banned or having their account disabled. Rather, the anonymous moderators limit the visibility of such users’ posts. Thus, invisibility becomes a form of punishment for not conforming to the established platform norms and identity. BNBR policies remain vague, giving participants little concrete guidance on what falls within or outside social norms of a heterogeneous Bengali community. Other participants likewise police the forums and can tag, upvote, downvote or report

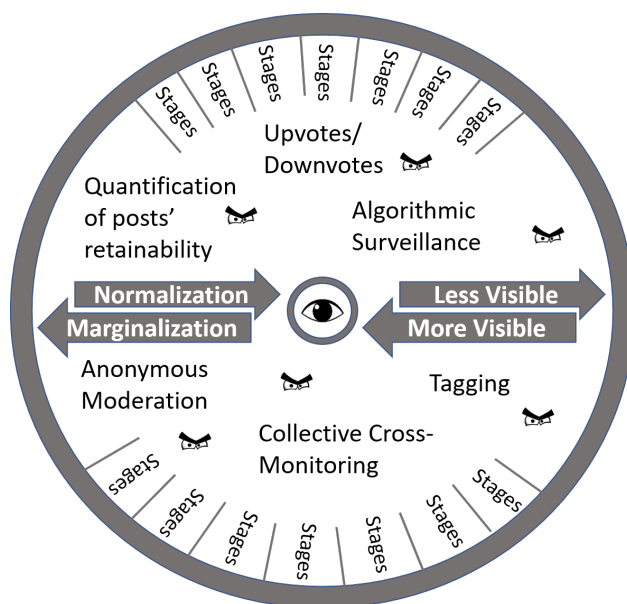


Figure 6.1: Panoptic Platform depiction of BnQuora.

posts as they see fit which creates a majoritarianism on this platform. The readings and voting practices by the community gets amplified by a recommendation algorithm. Popular posts gain more visibility while performances not matching the interest by the majority get pushed to the margins and become less and less visible. The majority gradually crowds out minority performances through this process of collective cross-monitoring, upvotes/downvotes and quantification of posts' retainability. In other words, the recommendation algorithm on Quora further accelerates these centrifugal and centripetal forces – bringing the normalized identities to a position of higher visibility while pushing the marginalized groups to positions of lesser visibility (see Figure 6.1).

The linguistic identity that initially brought together a range of discourse practices, temporal horizons, national and religious identities slowly purges minority performances or pushes them to the margins. Or, feeling the centrifugal push towards invisibility participants are given the choice to create their own marginal stages out of the majority's sight. Here, participants can freely exercise their preferred discourse practice, national and religious identity and interpret decolonization based on the temporal horizon they see fit. The platform does not become the arena for a shared

decolonization debate and narrative but one normalizing the majority identity and, in the process, marginalizing minority identities.

The data that we collected and analyzed from BnQuora in this chapter helped us understand the users' perception about the emerging platform. That they were able to join this platform and voice their frustrations about perceived moderation biases showed that they were not completely powerless. HCI research that looks through a postcolonial computing lens tends often to either overlook or exaggerate this agency of the users [329]. These agencies can be both constructive and destructive. For example, as the "stage" feature was introduced to BnQuora, the users could create and decide to join different stages according to their choices to engage in discussions with more similar-minded users. On the one hand, this created a more intimate space, potentially improving the potentiality of that online space for marginalized identities. On the other hand, certain groups felt confined within the boundaries of stages which further marginalizes those identities and increases the risk of segregation. With majority views dominating the major discussion forums, users with different perspectives get pushed into smaller echo-chambers. The exchange of perspective on the platform loses out. This not only fails Quora's objective of understanding each other better, but also increases the possibility of ideological polarization on the platform.

## Chapter 7

### Content Creation to Reimagine Transnational Communities

Nationalism creates a collective sense of identity—how individuals, despite their physical separation and limited personal interaction, see themselves as belonging to a large group [28] such as based on language, culture, religion, and even geopolitical borders. However, nationalism has a long and sordid history; a history where, when associated with modern nation-states, highlights a deep entanglement with the global history of colonialism [28, 111, 198]. Colonization exploited the local and Indigenous communities' resources and altered their political, economic, and cultural structures and identities [352], and later shaped local communities in terms of nation-states [111].

To attain the objectives of decolonization, colonized communities must conceptualize their native culture and identity outside the colonial influences exercised through cultural imposition and territorial divisions in the form of nation-states. While the restoration of native culture and identity is often used as an umbrella concept in the literature as the decolonial objective [513], it abstracts the myriad historical internal and external dislocations and transnational fractures colonization caused to people's collective identities. While nationalism embeds fragmentary ways to imagine their communities and collective identities within people, different sociopolitical structures and institutions, such as mainstream media and education systems, materially bolster those ideas in postcolonial nation-states. Acknowledging the social psychology and materiality of decolonization as a political struggle [198, 330, 565], in this chapter, we focus on decolonial discourse—discussions and practices that critically engage with and challenge colonial structures and ideologies.

As institutions in postcolonial nation-states continue to reify colonial legacies through promoting dominant narratives and hyper-nationalism [331, 472], the processes of decolonization involve rediscovering cultural heritage, mourning about history, exploring decolonial possibilities, committing to reforming practices, and taking concrete actions [330]. Especially considering decolonization's socio-psychological aspects [198, 199], as we explored in Chapter 5, recovering cultural identity, understanding history, and examining colonial influence in regional geopolitics, economic hurdles, and social injustices become contributing factors to the decolonial discourse, which often takes shape and manifests through technology like online platforms [152, 330]. Building on the research on how user-generated video-sharing online platforms support people in expressing and negotiating their cultural identities [116, 389] and participating in sociopolitical discussions [36, 39], we want to understand how sociotechnical systems like user-generated video-sharing platforms support decolonial discourse in transnational contexts given colonialism's complex impacts on nationalism.

To answer this question, in this chapter, through a qualitative study realized through semi-structured interviews with participants from Bangladesh, India, and Pakistan, we investigated YouTubers' motivations and strategies for video-mediated transnational decolonial discourse. Our findings reveal how communities imagined in relation to language, religion, and postcolonial nation-states and institutionalized nationalism through different sociopolitical, media, and educational systems motivate YouTubers in decolonial discourse on YouTube. Moreover, our study explains their strategies, such as political explainers, storytelling, and YouTube journalism, that contribute to decolonial discourse. Our findings explain the double bind of nationalism—historically contributing significantly to anticolonial movements and simultaneously isolating regional and local identities bearing colonial divisions—and how it manifests in decolonial discourses on YouTube.

## 7.1 Background

Building on the high-level literature review in Chapter 2, in this section, we will elaborate our discussion on collective identity, especially how people perceive communities through the lens of

nationalism and how colonialism has historically shaped its conceptualization. We will also discuss how postcolonial institutions shape hegemonic nationalist discourse and how people reimagine communities through technologies, such as video-mediated online platforms.

### **7.1.1 Nationalism as a Collective Identity**

One of the most prevalent ways people express and experience their collective identity is through nationalism. While nationalism is colloquially viewed as a political ideology, we draw on Benedict Anderson, whose seminal work exploring its origins [28] defines nations as limited communities containing people with the same interests and traits. According to Anderson, nationalism is a cultural system akin to religious beliefs, offering a sense of continuity in a contingent world. Using the concept of “imagined community”, he explains how people living in modern civilizations imagine connections to other citizens despite the impossibility of interacting with everyone in their society. He argued that nations were not the determinate products of given static sociological conditions such as language, race, or religion but rather imagined into existence despite the impossibility of interacting with everyone within a group. For example, he describes how Muslim communities in various geographic locations, even when they are unlikely to interact, share a sense of unity and camaraderie. He emphasized the sense of connectedness in the minds of the individuals. Prioritizing the role of language in fostering a sense of affinity, Anderson described the importance of institutions like the printing press through mass production, circulation, and standardization of printed materials, which he dubbed “print capitalism” in shaping and spreading shared identities and imagined communities, in turn developing modern nationalism. Drawing on the example of the dilemma of multilingual European empires, he argues that the idea of similarity in appearance, language, and practice is at the heart of nationalism-based collective identity. As communities undergo different experiences and exposure to different cultures and ideas, their collective identity can evolve and change over time.

### 7.1.2 How Colonialism Has Shaped Nationalism

One of the often normalized and invisible mechanisms that have shaped and continue to shape people's experiences and perceptions of nationalism is colonialism [62, 152]. While in the contemporary discourse of development and modernization, nationalism has been relegated as "a matter of ethnic politics," not too long ago, it was seen as "one of Europe's magnificent gifts to the rest of the world" [111]. Based on how they viewed nations and how people should be grouped, British colonizers proposed and, in some cases, utilized dichotomous schemes to divide various countries (e.g., two-nation theory in India-Pakistan, two-state solution in Israel-Palestine) [243], and did not consider the complex ways in which people see themselves or see themselves in relation to others.

Such partitions divided peoples into various nation-states based on Western ideals of nationalism. This is best explored by Partha Chatterjee, who, in his influential work "The Nation and Its Fragments" [111], criticized the European ideals of nationalism on which Benedict Anderson developed the foundation for imagined community. Here, Chatterjee argues that nationalism was a colonizing force imposed by "modular" forms of nationalism imposed by Western powers, questioning whether or not peoples who were colonized had anything "...left to imagine?". Modularity, in the context of nationalism, refers to the shared characteristics constituting imagined communities. The creative and powerful imaginations of nations in Asia and Africa, instead of being defined on their own terms—or creating modules of similarity as generated through their own sense of what it means to be a "collective"—were generated based on difference and the ideals of nationalism as imposed by Western powers. Using anticolonial nationalism in Bengal as evidence, Chatterjee demonstrated that while operating within the Western project of modernity, among the bilingual intelligentsia, the ideation of Bengali nationalism through evolving literary practices and cultural expression differentiated itself as "recognizably Indian." Though this linguistically grounded nationalism was a strong driving force during different stages of the independence movement [310], at

the end of British colonization, religion-based nationalism took priority over Bengali nationalism, leading to Bengal's partition.

This is best illustrated by how Western powers constructed nation-states through divide-and-rule policies. Divide-and-rule policies during the colonial period exacerbated religious animosity, resulting in significant communal violence towards the end of that era [431] and ultimately culminating in the formation of two nation-states, India and Pakistan, based on the two-nation theory [163]. Here, Western powers developed nation-states and subsequently modularized imagined communities based on external perceptions of religion mediated by the central premise that "...the Hindus and Muslims are two separate nations who cannot live together" [302]. The colonial approach to perceiving communities based on a monolithic idea of religion overlooking native sociocultural complexities, nuances, interconnections, values, traditions, and history, creates myriad internal and external fractures in the communities' self-perception. Hence, in postcolonial nation-states, repairing identity and "indigenizing the limbus"—the process of reinventing tradition, language, and culture face diverse conceptualizations of indigeneity and nationhood [552]. Multiple possibilities emerge for people to imagine communities across different dimensions, such as language (e.g., Bengali), religion (e.g., Hindu-Muslim), and post-partition nation-states (e.g., Indian-Pakistani-Bangladeshi), and negotiations among one's nationhood across various dimensions (e.g., language, religion, country) are often in flux. For example, in 1947, Muslim-majority East Bengal became part of Pakistan based on its shared religious identity. A few decades later, due to West Pakistan's linguistic-cultural imposition on East Bengal, Bengali nationalism started rising rapidly and replaced Pakistani nationalism to differentiate Bengali Muslims from non-Bengali Pakistani Muslims. Eventually, Bengali nationalism led to the formation of the nation-state Bangladesh in 1971 [493]. In post-independence Bangladesh, Madan studied "two faces of Bengali ethnicity": Muslim Bengali and Bengali Muslim [360], based on whichever identity people put forth. Similarly, the decades-long contemporary struggle of non-Bengali and non-Muslim *adivasi*<sup>1</sup> communities in Bangladesh highlights the nationalism-based tension across ethnolinguistic and religious differ-

---

<sup>1</sup> Heterogeneous tribal groups and ethnic minorities across the Indian subcontinent.

ences [308]. Such prioritization between multi-faceted identities is an individuated concept. In our study on Quora (Chapter 5), we found similar questions (e.g., “*Are you first a Muslim or a Bengali or a Bangladeshi?*” or “*Are you first a Hindu or a Bengali or an Indian?*”) to be major points of negotiation for identity in relation to their colonial past.

### 7.1.3 Institutions in Shaping Hegemonic Nationalist Discourse

In this chapter, we are more broadly interested in understanding why and how people engage in decolonial practices, which is a process wherein colonized territories and societies work to reclaim autonomy and independence from colonial powers [198, 330]. However, as previously described, colonization has created incredible complexity around people’s perceived imagined communities or nationhood in colonized and postcolonial societies, which can impact people’s opportunities or willingness to engage in decoloniality. While postcolonial nation-states like India, Pakistan, and Bangladesh mark a continuity in colonial legacies and institutions, they also embody a rupture from the colonial past, adding complexity to the decolonial discourse. This is further exacerbated by how coloniality shapes not only national identities but also sociocultural shifts, geopolitical processes and interests, and economic institutions within nation-states.

While establishing nation-states and independence from colonial rule is often seen as the end goal of decolonization, scholars have criticized this view as a myopic conceptualization of decolonial objectives [198, 330]. After the departure of colonial rule, oftentimes elites from the previously colonized communities, whom decolonial scholars described as “colonized intellectuals” [198], “interlocutors” [31], “bilingual intelligentsia” [111], occupy the helm of postcolonial nation-state’s political and administrative structures. These individuals, through their education and exposure to Western thought and adjacency with colonial rulers, while being members of the oppressed group, had absorbed the ideas, values, and cultural norms of the colonizers—what is known as colonial mentality [198, 247, 417]. Consequently, governance and other sociopolitical institutions in newly established nation-states often continue to perpetuate and reinforce colonial hierarchies (e.g., through forced integration of smaller ethnic minorities and religious majoritarianism) [434, 448].

Following the formation of a nation-state, reforming its social, political, and economic structures and practices in a way that reflects the culture and values of the diverse national communities should become decolonization's key objective [198].

Chatterjee argues anticolonial nationalism provides a formula for creating its own domain of sovereignty in political struggle within societies shaped by colonial influences, whether under the rule of foreign colonizers or their native interlocutors [111]. He discussed a fundamental feature of anticolonial nationalism in Asia and Africa [111] wherein it divides the social institutions and practices into two domains—the “outside” material domain of the economy, state-craft, science, and technology, and the “inner” spiritual domain bearing the “essential” marks of cultural identity. He emphasizes preserving the distinctiveness of culture that would guide careful consideration and reconfiguration of technological and material advancements of modernity to reflect previously colonially marginalized communities' values.

Yet, technological and material advancements continue to perpetuate anticolonial ideals and systems of power and privilege. Providing a historicist example, Chatterjee writes, “An entire institutional network of printing presses, publishing houses, newspapers, magazines, and literary societies is created ... outside the purview of the state and European missionaries, through which the new language is given shape. [They] came to think of its own language as belonging to that inner domain of cultural identity, from which the colonial intruder had to be kept out” [111]. Here, he underscores how anticolonial nationalism contributes to the local communities' cultural expression through media and literary institutions. However, in the construction of national identity during the colonial period, postcolonial scholars have critiqued elite groups establishing hegemony—socio-ideological control over norms, values, beliefs, and institutions [239], while disregarding the views of the subalterns—underserved individuals or groups who experience social, political, and economic marginalization within both colonial and postcolonial societies [248, 537]. Thus, it becomes important to reflect on “whose imagined communities” are reflected through institutionalized nationalism. In doing so, Bengali historians studied the political struggle between the elite financially

sound dominant caste Hindu landlords and the subaltern financially disadvantaged peasants from underprivileged caste Hindus and Muslims [111, 112]

Under the influence of a hegemonic nationalist agenda, besides the political structures of the postcolonial nation-states, institutions such as media and education systems continue to perpetuate a power hierarchy that leaves mass colonized populations out of shaping the socio-historical narratives. Scholars have found media representations to perpetuate colonial, prejudiced, and biased messaging [255]. For example, Indian and Pakistani media have been found to create narratives that alienate religious minorities and reify animosity between neighboring countries [365, 472]. Similarly, researchers also criticized how the education systems in this region sometimes propagate certain narratives that teach hyper-nationalism, politically charged historical narratives, and hatred toward neighboring countries [331, 615]. To challenge and subvert the hegemonic narratives and discourses shaped by colonial powers or elite cultures, decolonial scholarship, therefore, recommends decolonial discourse and practices—critical engagements and actions to shape public consciousness and societal narratives about continued colonial impacts on present-day societies. Decolonial activists have highlighted approaches to create counter-narratives, challenge colonial ideologies, amplify diverse voices, and promote critical engagement with media and inclusive pedagogical approaches and curriculum reforms that reflect accurate and inclusive histories, foster cultural pride, and promote a diverse understanding of the world [255, 330, 351, 554].

#### **7.1.4 Reimagining Communities through Decolonial Discourse on Online Platforms**

In the present digitally mediated era, this leads to the question of how digital media shapes these discourses. Decolonial discourse on online platforms is often mediated through and takes the form of critical conversations on historical and contemporary social, geopolitical, and economic issues and foregrounds diverse local cultural identities and perspectives [152]. Researchers have explored how technology can facilitate and hinder democratic processes, civic participation, and collective action [173, 210, 346, 541]. Looking at grassroots US political movements, Goshal et

al. explained how these sociotechnical systems exclude people based on racial, gender, and socioeconomic privileges and shape ideological hegemony [226, 227, 512]. Scrutinizing this notion of hegemony in the South Asian political landscape, a series of previous works in CHI, CSCW, and ICTD has looked at how religion [157, 428], caste [572, 573], social capital [508], and popularity [323] factor into political polarization and marginalization. Besides these empirical studies, a significant body of research focused on designing and evaluating systems that empower citizens. Examples of such endeavors include online civic platforms for deliberating democracy [369, 501], participatory budgeting as a part of open government and transparency [317, 318], and digital activism [340, 588].

While digital media platforms (e.g., Quora, YouTube) in many ways support marginalized communities' identity work—the process through which people construct, manage, present, and negotiate their identities [275], these platforms' censorship, copyright practices, and politics impede content creators' freedom and further impede marginalized communities' identity work [205, 208, 231]. Such duality is reflected in CSCW and HCI scholarship as researchers champion and critique the same platforms, respectively, for their support and hindrance in users' identity expression—in the same study [252, 329] or separate ones [24, 26, 149, 152]. This study is most centrally interested in how digital media contributes to people's engagement in decolonial discourse and identity work that serves as a reclamation project of their local and indigenous identities—a phenomenon that has been dubbed identity decolonization work [152]. Generally, we know from prior work that digital media can provide opportunities through which marginalized communities can push back against heteronormative societal logics by establishing a community to engage in identity work [187] as well as collective sensemaking during times of conflict [13, 367], amongst others. Our research builds on prior work focused on textual communications and social networks on online platforms (e.g., Reddit, Quora), bringing together a conceptual framework drawing on imagined communities and nationalism to explore people's sociopolitical engagement and identity decolonization work through video-based media.

Researchers have studied video-based platforms of various content lengths, origins, scales, and content types, such as the short-form video-sharing platforms Tiktok and Vine [53, 377, 518],

hyper-local platforms like Douyin and Kuaishou [116, 354], and international and generally longer video-hosting platform YouTube [389, 616], which is viewed through the lenses of CSCW and other adjacent fields like CHI, science and technology studies (STS), and media anthropology as a video-mediated platform that enables users to upload, watch, and interact with a wide array of user-generated and professional content [103]. Scholarships in these fields analyze YouTube's role in media ecosystems, business models, sociopolitical implications, technological infrastructure, and governance structures' impact on content production and distribution [95, 336, 358].

Studies on user-generated video-sharing online communities have highlighted these platforms' effectiveness in enacting their individuated and shared identities and connecting with others with common interests [39, 388, 616]. Askanius and colleagues deeply explored the use of YouTube videos for sociopolitical purposes [39, 570], various genres [36], evolution [37], and their convergence with mainstream media [38]. Focusing on the representation of cultural practices, Milliken and colleagues assess the role and contribution of user-generated online videos in creating online public sphere [388] and examine their potential and limitations [390], particularly in relation to regional identity [389]. Recent studies and media reports have highlighted how ethnic minorities in non-Western contexts use live streaming and video blogging (*vlogging*) to archive, showcase, and promote intangible cultural heritage activities for cultural sustainability [116, 142, 354].

Recently, researchers have studied how racial minorities in the US engage in political conversation in relation to colonial structures and build resilience to retain collective memories and revise colonial narratives using Reddit [173, 175]. Particularly in the Bengali context, we found in Chapters 5 and 6 how Bengali users collaboratively work toward reclaiming narrative agency for decolonizing their identity [152] and how sociotechnical platforms (e.g., Quora) reinforce colonial divisions and hierarchies, affecting users' participation and expression of cultural identity [149]. Those chapters highlighted users' perceived tension across various categorical identities derived from their conceptualized social groups or imagined communities and nationhood, such as the ones defined based on their religion and postcolonial nation-states [149, 152].

This dissertation investigates why and how previously colonized peoples from different nation-states are re-imagining their communities by engaging in decolonial discourse through video-mediated platforms. Here, we draw on Bengali scholar Chatterjee’s conceptualization of nationalism [111] and explore how decolonial discourse is shaped by national and collective identities, emerging from the convergence and differences of diverse aspects such as language, religion, and country and how a video-mediated platform like YouTube influences YouTubers’ strategies and content.

## 7.2 Methodological Details

This section describes how we identified and contacted potential participants—YouTubers who engage in video-mediated decolonial discourse. In addition to the demographic information of our participants, we provide details about how we interviewed them. Here, we also explain the process and various considerations in our data analysis.

### 7.2.1 Recruitment

Data for this study draws on semi-structured interviews with 15 content creators on YouTube living in Bangladesh, India, or Pakistan. Our eligibility criteria included: participants must be (1) 18 years or older, (2) a creator of YouTube content focused on Bangladesh, India, and/or Pakistan, and (3) had to be living in one of Bangladesh, India, and/or Pakistan. It is important to note that the researcher is a member of a minority group from Bangladesh and an avid YouTube user.

We identified participants for our study through a combination of purposive sampling [555] and snowball sampling [236]. Specifically, before recruiting participants, we spent time identifying potential participants we could contact to enroll in the study. Broadly conceived, our purposive sampling strategy included a combination of identifying and recruiting participants by (1) searching for content creators on YouTube and (2) recruiting from personal social networks. Prior works have highlighted that different linguistic, religious, and national identities are central to online decolonial discourse within the Bengali geocultural context [149, 152]. We searched for YouTube videos

using combinations of Bengali identity-related keywords, such as Bengali, Bangladesh, Bangladeshi, India, Indian, Pakistan, and Pakistani, as previously used by [149]. Though among the retrieved results, a large number of videos were Bengali movies, dramas, and music, since the focus of this study is discourse, we identified videos structured as dialogues and commentaries. Particularly, we identified channels operated from the region that frequently published videos about geographic and cultural identities. To ensure a more diverse sample, we also relied on YouTube's algorithmic recommendations for related videos and channels. This approach is adapted from prior work [150], allowing us to expand the diversity and scope of channels that were part of our larger sample. Moreover, we recruited from personal social networks. After identifying YouTube channels that regularly publish videos about Bengali culture, identity, and people, we considered the YouTubers of those channels as our potential participants.

To enroll participants in our study, we engaged in a multi-site endeavor. Due to the lack of a direct messaging feature within YouTube, we had to engage in myriad activities to contact and recruit participants. Whereas some YouTubers provided contact information in the description section of their channel, others did not. We focused on those YouTube channels where the creators provided a means for contacting them, such as through email or social media. For those channels that did not provide this information, we did not pursue them as it was a challenge to identify alternatives for communicating with them. We collected YouTubers' contact information from those descriptions, which included one or more of the following: email and Instagram, Facebook, and Twitter handle. First, we tried to contact them through email, which included a recruitment flyer describing the study's objectives, participants' eligibility, general information about the interview procedure, and researchers' contact information. Given people's varied email response behavior (e.g., rarely checking for emails), if we did not get a reply from a YouTuber within a week, we sent them a reminder email. If we could not contact a YouTuber through emails, we contacted them on social media from the researcher's profiles on corresponding platforms. We also contacted a few potential participants through real-life social networks. Following interviews, we engaged in snowball sampling, wherein we asked participants if they could recommend and connect us

with other potential participants. Recruiting and interviewing participants continued until we met theoretical saturation. We enrolled 15 participants in total. We summarize their demographic information like gender, country of nationality, age, religion, education, and occupation in Table 7.1.

Table 7.1: Demographic information of the participants

<b>ID</b>	<b>Gender</b>	<b>Country of nationality</b>	<b>Age</b>	<b>Religion</b>	<b>Education</b>	<b>Occupation</b>
P1	Male	India	20-24	Muslim	Bachelor’s	Engineer
P2	Male	India	35-40	Hindu	Master’s	Journalist
P3	Male	Pakistan	25-30	Muslim	Bachelor’s	Student
P4	Male	Bangladesh	25-30	Muslim	Master’s	Journalist
P5	Male	Bangladesh	30-34	Muslim	High school	Freelancer
P6	Female	India	30-34	<i>Did not disclose</i>	Bachelor’s	TV presenter
P7	Male	Bangladesh	30-34	Muslim	Bachelor’s	YouTuber
P8	Male	Bangladesh	30-34	Muslim	Master’s	Journalist
P9	Male	India	40-44	Hindu	Master’s	Govt. employee
P10	Male	Pakistan	25-30	Muslim	Bachelor’s	Engineer
P11	Female	Bangladesh	25-30	Muslim	Master’s	Job Aspirant
P12	Female	India	20-24	Hindu	Bachelor’s	Student
P13	Male	Pakistan	30-34	Muslim	Master’s	Engineer
P14	Male	India	20-24	Hindu	Master’s	Web-developer
P15	Female	India	20-24	Hindu	Master’s	Student

### 7.2.2 Interviews

Following the qualitative methodology outlined by Strauss and Corbin [547] and Yin [617], we conducted 15 in-depth semi-structured interviews between the Summer of 2020 and the Summer of 2022. Interviews lasted from 30 minutes to 1 hour and 53 minutes (averaging approximately 60 minutes). Given that all of our participants were international and physically distributed across different regions of Bangladesh, India, and Pakistan, interviews were conducted using the technology that was most comfortable and accessible to participants, including Zoom and telephone. In some cases, when the Internet connection was unstable, we switched from Zoom to telephone. Participation in this study was voluntary, meaning participants did not receive remuneration for their participation. Since we initially intended for the interviews to last only one hour, we alerted

participants when we reached that time limit. However, all of them wanted to continue beyond the pre-decided duration. Prior to initiating any given interview, we read an oral consent form to the participants. All participants provided verbal consent and also consented to the interviews being recorded. Importantly, the researcher is a native Bengali speaker with bilingual proficiency in English and a working oral proficiency in Hindi/Urdu<sup>2</sup>. He conducted all the interviews in Bengali, English, or Hindi/Urdu based on participant preferences.

Interviews were designed as life histories [599], seeking to understand and locate people’s video content creation within their long-term life experiences. We initiated interviews with demographic questions, followed by questions seeking to understand how their family and life histories have shaped their perception of their sociohistorical perspectives. We then asked participants about their motivations and online and offline experiences for making videos on Bengali culture, history, and society, followed by questions about their video-making practices. Each interview, on average, was approximately an hour long. We transcribed the interview recordings and translated the non-English interviews into English. We translated the Bengali interviews, while a researcher assistant, a native Hindi speaker, did the same for the Hindi/Urdu interviews. All transcripts were anonymized and de-identified before analysis.

### 7.2.3 Data Analysis

We analyzed our data using an inductive, grounded theory [547] inspired approach, commonly used in qualitative HCI and social computing studies [149, 269, 569]. We used qualitative data analysis software MaxQDA Plus to code the interview transcripts. Following Strauss and Corbin’s guidelines [547], we analyzed our data in three phases. In the open coding phase, we iteratively reviewed the interview transcripts and identified the repeatedly appearing abstract concepts, events, and interactions. Some examples of open codes emerging in this phase are: *“ancestral ties and cultural connection”*, *“childhood exposure to communal violence and political turmoil”*, and

---

<sup>2</sup> As spoken languages, Hindi and Urdu are mutually intelligible, to the point that they are sometimes considered dialects or registers of a single spoken language known as Hindi-Urdu or Hindustani [81, 129].

*“living memories of partition and migration as refugees”*. We associated quotes from the participants with the corresponding open codes. We then collaboratively engaged in axial coding, where we combined the open codes to create higher conceptual themes. For example, aforementioned open codes were merged to create the axial code *“personal ties and lived experiences”*. Finally, we identified the relationships between the axial codes in the selective coding phase, resulting in the themes, i.e., motivations and strategies, we present in this chapter. Since interview data is contextual, our reflexive and interpretivist grounded theory-based data analysis [494, 547] does not require us to calculate an inter-coder reliability score [374]. Moreover, it is important to note that while our work stayed true to interpretivist tradition, that means the analysis was not only “about” the phenomenon under study but equally, implicitly or explicitly, “about” the perspective of the researchers [29]. Due to their embeddedness within colonially shaped sociocultural contexts that motivated their research, we acknowledge that their work, which focuses on conversations about colonial histories, is inherently political. The researcher’s membership in intersectionally marginalized communities, lived experiences in colonially impacted societies, and past experience in critical (e.g., decolonial, postcolonial) computing research motivated this study of colonially marginalized communities’ practices around technology. Thus, in addition to the context of the work and its focus on conversations about colonial histories, which are inherently political, our unique demographic and scholarly backgrounds have shaped the decolonial interpretations of the data that are presented in this work.

### **7.3 Results**

During our interviews, our participants openly shared their family history and personal lived experiences. Throughout these conversations, we discovered that the partition of Bengal into different nation-states led to isolation and communication gaps among the Bengali people. As a result, our informants described “using their imagination” to develop perceptions of their imagined Bengali community. Our participants discussed how their everyday experiences motivated them to

delve deeper into their culture and history. For example, Participant P12 explained her perception of transnational collective Bengali identity, driven by her family history and ancestral ties:

*We often talk about Bangladesh in our house. Ancestors of both my parents are from Bangladesh. My father grew up in [a Bengali-speaking Indian state]. There is a place called X where he lived with my grandmother and his five other siblings. They struggled a lot after migrating from Bangladesh to India. My mother grew up in [another Bengali-speaking Indian state]. Because the Bengali people have a very strong attachment to Bangladesh, I am interested in Bangladesh.* (P12, female, India)

Participants discussed how their multi-faceted linguistic dialects and internalized cultural customs strengthened their connection and attachment to people in neighboring countries beyond physical borders. This connection to their Bengali identities and communities that were dislocated due to historical colonialism has led to two broad forms of decolonial discourse on YouTube that fall in line with Chatterjee's assertion that nationalism drives anticolonial practices in both the "inner" domain of cultural identity and the "outside" domain of institutionalization [111]. We explore the motivations of YouTubers' video-mediated decolonial discourses as characterized by their experiences with (1) nationalism's impact on the cultural community, the "inner" domain of cultural identity, and (2) the "outside" domain of nationalism defined through institutionalization, where their strategies include making different kinds of videos, such as travel vlogs, social interviews, reaction and journalistic videos, political explainers, and satires. We also elaborate on how each kind of video contributes to the Bengali decolonial discourse on YouTube.

### **7.3.1 Nationalism's Impact on Cultural Community: The "Inner" Domain of Cultural Identity**

Conceptualizing and committing to a vision of native culture and national identity are vital to decolonization [198, 330]. We examined YouTubers' reflections on how the colonial history of

Bengal has socio-psychologically conditioned people in this region to perceive the “inner” domain of cultural identity—their collective native identity—and how they strategize their video-making to bridge isolated and disintegrated conceptualizations of Bengali culture and identity.

In the sections that follow, we first describe people’s motivations for producing YouTube discourses centered around their collective identities. We then describe the decolonial strategies they employ as part of their identity decolonization work.

### 7.3.1.1 Motivations

Our findings highlight how participants use language-based culture, religion, and post-partition nation-states as threads for identifying fractures. By negotiating collective identities, they foster affinity beyond national borders. We elaborate on how their transnational conceptualization of communities and interaction with their viewers affect their content creation and engagement. We present their motivations through the following themes: (a) language-based cultural similarities and diversity, (b) religion driving reconciliation and empathy, and (c) negotiating identities based on postcolonial partitions.

**Re-Imagining Community Around Language-based Cultural Similarities and Diversity** Bengali YouTubers highlighted celebrating the similarities and diverse ways in which people express and practice local ethnic identities and associated linguistic and cultural practices in Bangladesh and India as their motivation behind making videos on YouTube about their culture.

Despite their community being displaced across the fifth longest and one of the most dangerous international borders in the world [97], the YouTubers in these two countries were motivated by the opportunities to re-imagine their community based on their language; they wished to celebrate and highlight their linguistic identity as a common feature of their displaced community. Participant P15 described her rationale and strategy to foreground her Bengali identity through her YouTube channel:

*We are Bengali, and we speak in Bengali. Since the national language of Bangladesh is Bengali, they [Bangladeshis] would be able to connect with and relate to our videos. ... We always try to prioritize the Bengali language in our channel. So, we named our channel X [explains the name's etymological relation to Bengali] (P15, female, India)*

The associated cultures and customs as shaped through linguistic norms also inspire the YouTubers to connect with people in “*Opar Bangla*”<sup>3</sup>. These YouTubers imagine themselves and their audiences as part of a broader community of Bengali identity around the linguistic dimension and beyond geographic boundaries. Examples include Bengali norms of addressing each other even when they have never met as “*dada*” (দাদা: brother) or “*didi*” (দিদি: sister) that foster closeness with their audiences. Participant P9 described how his perspective drives his intent to be on YouTube:

*We became two different countries, two different peoples. In fact, I did not know much about them [Bangladeshi Bengali audiences]. I became closer and more cordial with them when I got to know them better. I was fascinated by their manner and my conversations with them. ... The space of cordiality widened among us. (P9, male, India)*

The diversity of dialects introduced by several migrations between two sides of Bengal creates a probability for some Indian Bengali (e.g., P12) and some Bangladeshi Bengali (e.g., P7) YouTubers talked about having similar dialects as Bangladeshi Bengali and Indian Bengali audiences, respectively. Those participants think that this builds a perception of belonging to a further closer group based on similar dialects (e.g., *Bangal* and *Ghoti*<sup>4</sup>), within the broader language-

<sup>3</sup> Given the colonial history of partition of Bengal, Bengalis in Bangladesh (then East Bengal) and Indian state of West Bengal refer to their side of Bengal as “*epar Bangla*” (এপাড় বাংলা: “this side of Bengal”) and the part of Bengal on the other side of the Bangladesh-India border as “*opar Bangla*” (ওপাড় বাংলা: “the other side of Bengal”). These phrases are usually used in cultural and literary discussions [47, 113].

<sup>4</sup> Bangal and Ghoti are distinguishable variations of the Bengali language spoken by people generally in the Eastern and Western region, respectively [149].

based imagined community, motivating them to create content on diverse regional linguistic and folk cultures.

**Towards Religious Reconciliation and Empathy: Re-Imagining the Role of Religion in Community** Our participants acknowledged how religion had shaped their sense of community through colonization. They were motivated by the opportunity to reimagine the role of religion from one that created division to one that allows for reconciliation and empathy-building.

All our participants from Pakistan (P3, P10, P13) and some from Bangladesh (P4, P5) who belong to the Islamic faith have highlighted their religion as a strong aspect of their identity and see that as a connecting thread among people in transnational contexts beyond borders. To describe how his religious value motivates him to cooperate with other Muslims, P13 commented,

*As Muslims, we need to help each other. As our Prophet (PBUH) said that a Muslim is a brother of another Muslim. ... The question always comes up “What did you Muslims do in 1971?” ... There are always clashes between brothers on some issues. But that doesn’t mean you should get separated from each other and never connect.*

(P13, male, Pakistan)

The quote from this Pakistani YouTuber essentially uses religion as the basis for defining communities. He also promptly responded to a common critique of religion-based nationalism in the context of Pakistan’s oppression of Bengalis despite common religious identity and the genocide of Bengalis in 1971. We found similar sentiments among other Pakistani participants who make videos about Bengali culture. Their perspectives echo prior scholarships on Muslim identity and how it differentiates between believers (Muslims) and non-believers (non-Muslims) as a gatekeeping mechanism for imagining communities [20, 28]. Based on the Muslim Brotherhood, a strong willingness for reconciliation between Bangladesh and Pakistan was visible in their conversations. The same focus on harmony and cooperation is also present in these YouTubers’ videos, especially in cases of easier visa processing and increased financial cooperation.

Some Bangladeshi and Pakistani YouTubers (P4, P11, P13) highlighted how their common belief in Islam establishes a sense of community with Muslims within Indian borders. For example, Participant P4 shared his experience of visiting Kashmir, which is one of the Muslim-majority Indian union territories [543], and interacting with people there:

*When [Kashmiri people at the hotel] heard that we went there from Bangladesh and we were Muslims, they were saying that they were also Muslims, and so, we were brothers. They are hospitable and sweet talkers.* (P4, male, Bangladesh)

Similarly, some Indian YouTubers (P2, P12) expressed their sense of affinity with Bangladeshi Hindus based on their religious identity. In these cases, the religious similarities made the YouTubers compassionate and empathize with the struggles of people of the same religious identities in different countries and motivated the YouTubers to focus on religious minorities like Bangladeshi Hindus and Indian Muslims, for example, resulting in Bangladeshi and Pakistani YouTubers talking about Indian Muslims' marginalization and Indian YouTubers sympathizing with Bangladeshi Hindus during the times of religious violence. In some cases, this translated to empathizing with minorities in their own countries. For example, P2, an Indian Hindu YouTuber, described his effort to highlight the injustice to an Indian Muslim woman despite many likely being indifferent to it:

*We will release an episode on [a victim] today. I know it won't do even half a million [views] because who wants to know about the 2002 Gujarat riots and the rape of a Muslim woman there.* (P2, male, India)

As religion shapes participants' imagined communities, their videos bring forth minorities' experience with regional geopolitics (e.g., Kashmir), the rise of religious extremism, violence, and minority persecution—transgenerational issues tracing back to colonial divide-and-rule policies.

**Re-Imagining the Postcolonial Partition: Towards Understanding Geopolitical Neighbors** In contrast to the participants we mentioned so far, whose sense of imagined community

motivates their engagement in decolonial discourse, some YouTubers were motivated to produce videos that continued to perpetuate the colonial logic of partition that divided the Bengali people. They persist in the nationalist agenda that continues to impose a Western modularization of nationalism based on their specific nation-states. For example, Bangladeshi YouTubers often motivate their video-making practices from the point of Bangladeshi nationalism. Participant P4 described how Bangladeshi nationalism served to motivate his video production on YouTube:

*Bangladeshis watch lots of videos about what foreigners know or think about Bangladesh. This makes it clear that nationalism is strong among Bangladeshis. If someone praises Bangladesh, they become happy and listen to them attentively.*  
(P4, male, Bangladesh)

Similarly, a minority of Pakistani (P10, P13) and Indian (P14, P15) YouTubers described an impulse to propagate Pakistani and Indian nationalism. These YouTubers were deeply motivated by opportunities to imagine exclusive communities defined based on modern nation-states. These YouTube channels were shaped by hyper-nationalist views that harbor collective hatred towards individuals from other countries. Participant P4, the same individual as above, explained his perspective on post-independence Bangladeshi nationalism:

*Because of our history of the liberation war, whenever we talk to a Pakistani, the response your brain automatically gives you is that they are our enemy, they oppressed us, they dishonored our mothers and sisters, and they killed our fathers and brothers. This reality will shake you at the very beginning.* (P4, male, Bangladesh)

Here, the Bangladeshi Participant P4, who empathized with Kashmiri Muslims in India based on religion, is prioritizing his nation-state-based Bangladeshi identity over religious identity, focusing on the country's history of the liberation war against Pakistan. This exercise of nationalism, as

Chatterjee argued [111], leads the YouTubers to imagine communities based on differences. As an example of that, Bangladeshi nationalism is often defined through actively distancing themselves from Pakistani identity. In contrast, the postcolonial partition of the subcontinent in 1947, where Bangladesh used to be East Pakistan, influenced the Pakistani YouTubers' (P10, P13) imagination of a community. Participant P10 described his view:

*Whenever I hear the word Bangladesh, the first thought that comes to my mind is that it was a small part of us, known as East Pakistan, which later got separated. So, thinking about it now, it would be great if we were still together.* (P10, male, Pakistan)

From the angle of Pakistani nationalism, this participant mourns about Pakistan's history with Bangladesh, which Laenui defines as an important step in the decolonial process [330]. Echoing South Asian scholars' argument that Pakistani nationalism builds upon collective Muslim identity that separates itself from other religious communities [20, 163], the Pakistani YouTubers strongly tied their sense of nationalism to their strong remembrance of the two-nation theory. Contemporary political relations, interests, and conflicts and the rise of far-right exclusionary nationalism like Hindutva in India [406] and Islamism in Bangladesh [457], which are distinct from anti-colonial nationalism, also influence people's perception. Besides making YouTube empathize with minorities in other countries, as we previously discussed in section 7.3.1.1, some participants' use of negative phrases like "enemy" and "conspirator" to describe neighboring countries demonstrated how those factors accentuate the process of othering.

However, the majority of our participants highlighted how their motivations were guided by anticolonial logic and a decolonial impulse to re-imagine their post-partition relationships with their neighbors by seeking to understand their geopolitical neighbors' knowledge and perspectives about them. For example, participant P11 evaluated her audience's curiosity about other nation-states as an opportunity to develop mutual cultural and historical understanding:

*Especially when an Indian or a Pakistani person talks about Bangladesh's history and culture or what they know, for example, [about] what happened during the liberation war, or about our linguistic tradition or views on how Bangladesh is developing, people become interested.* (P11, female, Bangladesh)

Other YouTubers were motivated to portray their countries in a more positive light and explain their perspectives to their geopolitical neighbors. Given the negative attitudes from neighboring countries like Bangladesh and India toward Pakistan, some Pakistani YouTubers use the platform to advocate for their country. Participant P13 to describe his motivations, said,

*Every person is an ambassador of his country. ... Even South Asian countries ... have negative thoughts about Pakistan. And we were wondering why so. We are the people that others have negative thoughts about. So, we thought, ... we would promote our culture, our values, and a positive image of our country.* (P13, male, Pakistan)

Some of our participants (e.g., P11, as quoted earlier) believed that promoting their countries to a broader regional and international audience serves to manage how others see or feel about their collective identity, which makes their videos more likely to be popular among audiences from their countries. This pragmatic potential to gain popularity motivates YouTubers to make videos about their national identity and local cultures. While the decolonial discourse and practices should embody the “essential” marks of cultural identity due to colonial history and partitions, the Bengali culture and collective identity have been fragmented in the imaginaries of the people in this region.

### **7.3.1.2 Decolonial Strategies and Re-Imagining the “Inner” Domain of Cultural Identity**

To realize their stated motivations centered around revising existing linguistic, religious, and partition-based collective understandings, YouTubers adopt decolonial strategies to re-imagine

the “inner” domain of cultural identity. Particularly, YouTubers were working to re-imagine and establish a new collective Bengali identity on their own terms. The strategies employed include (a) fostering transnational cultural understanding through reaction videos and social interviews and (b) decolonial storytelling using travel video blogs at historical sites and cultural festivals as probes.

**Reaction Videos and Social Interviews to Foster Interconnected Cultural Understanding** Our participants often make videos about cultural artifacts and practices of their neighboring countries to foster an interconnected regional cultural identity, which are often structured as reaction videos, social interviews, or public reactions. They make videos showing their emotional reactions while watching entertainment artifacts like television series episodes, film trailers, music videos, and short documentaries. These videos are colloquially called “reaction videos” and are numerous and popular on video hosting services such as YouTube [602]. Through these videos, our participants try to identify and highlight intercultural similarities among Bengali communities’ cultures in different countries. As discussed earlier, they often focus on language, religion, or postcolonial partitions as axes while finding similarities. For example, participant P3 said:

*My videos that became popular were mostly videos about the Islamic scholar [name]. He is a famous scholar of Bangladesh. I get a lot of views on those videos. Many Pakistani and Bangladeshi people like those videos.* (P3, male, Pakistan)

Here, the participant described his strategy to understand his Bangladeshi viewers’ cultural practices using religion as a lens. Given the cruciality of religion in the region, our participants talked about accommodating different religions’ various influences within their culture instead of the “thoughtless adoption of European customs”. To do so, they made videos on Muslim festivals Eid-ul-Fitr and Eid-ul-Adha, Hindu festivals Diwali and Holi, and common Bengali festivals like “*Boshonto boron*” (বসন্ত বরণ: welcoming the Spring season). In other videos, some YouTubers (P13, P14, P15) have also made such reaction videos about historical figures and events (e.g., the language movement of 1952), popular personalities, cultural topics, and artifacts. Since assessing

the potential popularity of a topic in a different country can be difficult, our participants check out trending topics before making videos. Participant P1 explained his process as follows:

*I will go to the internet, and first, I will search for topics that are trending and suitable for my channel. I am now preparing a video about Bangladesh, and I am writing the script. A week ago, I searched on Google for “trending topics in Bangladesh.”* (P1, male, India)

Another type of video several of our participants make to foster a broader understanding of Bengali culture among people from transnational contexts is structured as social interviews. In these videos, YouTubers go to public places and ask random people about their knowledge and perceptions of people in neighboring countries. Our participants used phrases like “social experiment” or “public reaction” to describe this genre of videos. Participant P8 described the questions he asked while making some past “public reaction” videos:

*When we made videos in Kolkata or India, we asked, for example, what the people in Kolkata know about Bangladesh. Who are real Bengalis—the people in Bangladesh or the people in West Bengal [, India]? Or what do the people in Pakistan know about Bangladesh cricket?* (P8, male, Bangladesh)

To assess how much the people in their countries know about their neighboring countries, our participants use various physical artifacts (e.g., other countries’ flags and currencies). Participant P10, shared his strategies for “social experiments”:

*I made basic videos, such as videos on Bangladesh’s currency or flag. I would show some countries’ flags to people and ask them which was Bangladesh’s flag. ... I gradually went deeper and made more insightful videos about Bangladesh.* (P10, male, Pakistan)

Like reaction videos on entertainment artifacts, our participants emphasize finding similarities through these social interviews. Interestingly, the sport cricket as a postcolonial influence [361] excites people and serves as a connecting point for Bangladesh, India, and Pakistan. Their common food culture also brings people from these countries together. Participant P15 mentioned a recent video, in which she focused on the similar food cultures and ambiances in Bangladesh and India:

*We are making another video today on Bangladeshi street foods. We make videos about Dhaka. We shot a video yesterday at Purbachal<sup>5</sup>. That place seemed similar to our new town in Kolkata.* (P15, female, India)

Here, the participant describes how she perceived two neighborhoods, one Bangladeshi and another Indian, as being similar based on comparable availability and popularity of street food. Gathering at street corners over tea or street food to discuss intellectual matters or gossip among intimate friends, which postcolonial scholar Dipesh Chakrabarty dubbed as “*adda*” [107], in his words, is “something quintessentially Bengali, ... an indispensable part of the Bengali character” [107]. Similar to prior work [367] that found videos highlighting people’s everyday practices to be helpful in humanizing people in different social contexts, through these videos highlighting cultural practices and artifacts, our participants aim to facilitate a better cultural understanding by disproving myths and stereotypes about Bengali communities across borders and religions, shaping a more harmonious attitude toward each other.

**Decolonial Storytelling through Travel Vlogs to Highlight Local History and Culture** While recent scholarship in CSCW and CHI has discussed the role of video blogs (*vlogs*) and live streamings in cultural sustainability and representation of intangible cultural heritages, rituals, and practices [116, 354], our participants prioritize historical perspectives besides cultural festivals in their vlogs. Those who make vlogs (P9, P14, P15) often choose to structure those as travel vlogs—a particular type of vlog where they showcase and self-document their journeys to

---

<sup>5</sup> Among others, this place has recently been hyped for many food options that are good for groups, have all-you-can-eat options, or are warm and cozy.

places of Bengali historical importance. For example, participant P9 described one of his popular vlogs:

*Initially, I tried to present my videos like travel vlogs where I would show a new place to people on YouTube. In the early days of my channel, in a video, I showed indigo trees. I read a lot about the indigo plantations in Bengal by the British. But how many of us even recognize indigo trees? General people do not even know that there are a lot of indigo trees around us. I used that indigo tree as a “tool” to make one of the initial videos about British colonial rule—how they oppressed and forced the farmers to cultivate indigo. I got over 1.5 million views on that video. (P9, male, India)*

Here, the participant explained his attempt to highlight the Indigo plantation, which was a significant factor in the historic devastation of both East and West Bengal’s agriculture, yet is a globally lesser-highlighted aspect of British colonization in the Indian subcontinent. In doing so, he used an everyday material probe—an object inspiring open-ended and evocative activities like story-telling [611]. Such probing invites viewers to revitalize their historical sense and memory of the British colonial period, which is a critical component of decolonization discourse online [152]. As the history of common glory and, more importantly, that of common sufferings brings a nation together [453], these vlogs can evoke memories of past glories and sufferings among Bengali people, now living in different nation-states, leading them toward bolstered and interconnected Bengali identity. Many of our participants were born and raised in historic towns and villages—a fact in which our participants took pride. They often make these travel vlogs in sites neighboring their living places, with numerous instances of traveling further, such as major tourist attractions. Their focus on local historical sites underscores the deep-rooted nature of colonial influences, makes it relatable in people’s everyday experience, and contributes to constructing historical narratives from local points of view. In Participant P9’s words:

*Through YouTube, I can show all the nearby places and explain their history to people. I started from the often comparatively ignored places and their history.* (P9, male, India)

These local perspectives complement people’s historical understanding and add nuances to the native narratives of colonization in an attempt to foster a new and re-imagined Bengali collective identity. To surface a sense of historical reminiscence, our participants (e.g., P9, P14) make travel vlogs at old temples and mosques. As an alternate strategy to feature such Bengali cultural festivals, some of our participants also make travel vlogs at various iconic places (e.g., Rabindra Bharati University<sup>6</sup>). They believe that these physical explorations, self-documentations, and sharing of experiences serve as points of historical and cultural reflection and realization for themselves and their viewers. As probes (e.g., place, everyday objects) are effective in fostering decolonial storytelling about native and indigenous cultures [110, 467], they described YouTube as a more engaging and effective medium than textual blogs and encyclopedias for historical and cultural storytelling.

### **7.3.2 Nationalism Defined through Institutions and Establishments: The “Outside” Domain of Institutionalization**

Our participants discussed the nationalist agenda perpetuated by different institutions in their postcolonial nation-states. As institutionalized nationalism reinstates colonial values and reflects hegemonic perspectives, they also described their motivations and decolonial strategies, using video creation as a means to resist hierarchies and power imbalances.

#### **7.3.2.1 Motivations**

Since various institutions in modern nation-states perpetuate colonial legacies, our participants were motivated by the opportunity to re-imagine the institutions that have shaped their

---

<sup>6</sup> Rabindranath Tagore is a Noble laureate and famous Bengali poet. Rabindra Bharati University is a public university located at his family home, Jorasanko Thakur Bari, in Kolkata, India, founded to commemorate Tagore’s birth centenary.

collective imaginaries of what it means to be a community. We elaborate on this through the following motivations, including (1) identifying and resisting colonial impulses within sociopolitical structures, (2) revising education systems' fragmentary and selective historical narratives, and (3) re-imagining transnational representations in mainstream media. YouTubers elaborated how their observations of these institutions falling short of constructing comprehensive historical narratives motivated them to initiate decolonial discourse.

**Identifying and Resisting Colonial Impulses within Sociopolitical Structures** Chatterjee argues that through colonization, nationalism in the Indian subcontinent materializes through differentiation [111]. This differentiation wherein nation-states are weaponized against one another is reinforced by existing sociopolitical institutional structures, including politicians, numerous armed conflicts at borders, and adversarial exchanges at international venues are manifestations of nationalism through institutions [218, 230]. To address these transgenerational colonial impacts [152], our participants are motivated by opportunities to reimagine regional geopolitics and find ways to highlight the public desire for harmony and cooperation. For example, participant P10 said:

*The more fights there are between the two [countries: India and Pakistan], the more problems it will create for the respective countries. Money will also not be wasted on such wars, which can otherwise be used for better purposes, and people in both countries will be happier. ... for people in Pakistan and India, 70-75% people are against any kind of violence and want to live in peace. Both countries will only grow when 99-100% of people think wars should not happen. (P10, male, Pakistan)*

Here, he criticized the India-Pakistan relationship and explained how his country's political narrative differs from most people's perspectives. As regional political narratives, contemporary events of communal extremism, intolerance, and the rise of religious majoritarianism carry on colonial legacies of religion-based partition, Participant P2 shared how institutionalized religion in regional politics affected his childhood memories and motivated his discourse acts on YouTube:

*My earliest political recollection was the demolition of Babri Masjid<sup>7</sup> when I was 12 years old. I remember the curfew that we lived under for two weeks at least. Having seen the debate around it since school. I have always been interested in politics since school, not being a politician, but reporting and talking about the politics of it.*

(P2, male, India)

Similar to P2, whose childhood experiences motivated him to build political awareness, two of our Pakistani participants talked about the country's internal politics around Bengali people and this diaspora community's nationality and rights being denied.

Besides discussing these colonially designed social injustices, participants (P1, P2, P13) were further motivated by opportunities to examine how the institutionalization of coloniality has continued to shape their everyday experiences and existence, which are crucial decolonial practices [198, 330]. They purposefully made videos about economic struggles, reformation, and development to underscore the ongoing impacts of coloniality on relationships with their neighbors. For example, participant P1 described his interest in identifying economic impediments:

*Why are we so lagging behind? Why is there so much poverty in our countries? Why are they in so much economic distress? Why is there so much corruption? We have to look into all those and find common ground so that our people can prosper.*

(P1, male, India)

In doing so, YouTubers identified colonialism's continuing impacts, such as adversarial geopolitics, social injustices, and economic hurdles in subcontinental establishments, which they deem to be the prerequisite to finding ways to remain proactive in addressing those issues. They also scrutinized how colonial values and structures are reinforced through education systems and media.

### **Revising the Education Systems' Fragmentary and Selective Historical Narratives**

Like Chatterjee highlighted the role of print media and literary works in constructing historical

---

<sup>7</sup> A long-standing dispute over a religious site in Ayodhya, India, involving competing claims between Hindus and Muslims regarding the construction of a mosque or a temple at the location.

narratives and the idea of nations [111], we found that most of our participants were motivated to produce YouTube content to revise the narratives being produced by educational institutions. Informants described the role of textbooks and historical narratives featured in their national education system as an important factor in making them interested in Bengali sociohistoric backgrounds and motivating them to make videos in this space. Participant P11 described her perception of how the country's educational institutions construct a narrative about Bengali people's history:

*From fifth to tenth grade, our textbooks taught us the history of the Indian sub-continent in great detail. By the time we are in tenth grade, we are aware of our history, from ancient times to the history of our liberation war.* (P11, female, Bangladesh)

Though our participants' history education has been a strong motivator for their decolonial discourse on YouTube, they complained about cherry-picking in textbooks' historical narratives and thought that their history education should have highlighted a broader set of historical figures. The politics around history education was highlighted strongly by Pakistani YouTubers. For example, P3 explained how textbooks avoid talking about its history with Bangladesh's independence:

*Pakistani textbooks do not have any chapters that mention Bangladesh. The 1947 division of India and Pakistan is mentioned, and all the other things are mentioned as well, but there is nothing mentioned about anything that happened in 1971 [Bangladesh's liberation from Pakistan]. ... The people are unaware of the history.* (P3, male, Pakistan)

Another Pakistani participant, P13, described how the Pakistani educational system describes the Bengali liberation movement as a result of being "played by the Indians". This insight aligns with prior works on Pakistani education systems propagating animosity with neighboring countries as a way to strengthen nationalism [331]. Our participants discussed varied selective attention and

political efforts to shape accounts of historical events through education systems in different countries, curriculum boards, and partisan administrations. Their perspectives reaffirm Chatterjee's observation of Hindus and Muslims in the subcontinent viewing historical narratives by focusing on different timeframes and historical characters [111]. For example, our Pakistani participants mentioned their education system's exclusive focus on Muslim rule in India while overlooking non-Muslim freedom fighters' contributions. Similarly, our Indian participants expressed concerns about the recent erasures of Muslim historical figures from textbooks under the rise of right-wing politics. Participant P2 expressed disappointment at how historical narratives and communal politics have influenced each other in recent years:

*It is tragic how politics has been shaped. People have been harking back and trying to look for revenge for what happened in the 16th century. ... Maybe they don't blame the British as much as the Mughals, and we know why. The looting of the British was much, much more and much more catastrophic than even the Mughals. But the focus is on the Mughals because that's going to get you ahead in political points.*

(P2, male, India)

If fragmentary historical narratives institutionalized through education systems continue to condition future generations with exclusionary nationalist views, it would impede the sociopolitical structures of postcolonial nation-states from reforming and reflecting the masses' dreams that served as the basis of anti-colonial movements. Therefore, our participants make videos about regional history to complement and resist institutionally constructed fragmented narratives.

**Challenging Transnational Representations in Mainstream Media** Our participants from Bangladesh and India, imagining communities based on their shared linguistic identity, as discussed in section 7.3.1.1, were motivated to make YouTube content by opportunities to challenge transnational representations of Bengali identity by mainstream media institutions. Indian Bengali YouTubers (e.g., P12, P14, P15) discussed how mainstream media's representation of transnational Bengali culture contributed to fostering and sustaining that sense of affinity in them. They

talked about recent collaborations among media personalities, the availability of content from “*Opar Bangla*” on Over-the-top (OTT) platforms, and their contribution to representing cultural practices and underscoring similarities. Participant P12 talked about mainstream media and how it promotes harmony with the neighboring country:

*Indian media talks about Bangladesh. Recently, a singer from Bangladesh came to [a music talent show on Indian Bengali TV] and became the first runner-up. It's not that we are not in touch with Bangladesh at all.* (P12, female, India)

While our Indian participants viewed Bangladeshi culture's representation in their mainstream media as a thread of closeness and imagining a broader Bengali community, our Bangladeshi participants critiqued those as “caricatures” of and disconnected from contemporary Bangladeshi societies. Thus, our participants were motivated to produce YouTube content that challenged these caricatures and moved towards characterizations that they deemed appropriate and accurate. For example, participants criticized the portrayal of Bangladeshi characters by Indian-Bengali actors in various movies and TV series, as their ways of speaking did not accurately capture any Bangladeshi dialects and accents. Therefore, our participants deem it important to uphold grassroots practices from diverse national contexts instead of curated and stereotyped portrayals.

YouTubers have also discussed how hegemony and prejudices impact transnational media representation. Pakistani YouTubers (e.g., P3, P13) discussed how their mainstream media talks negatively about Indian people, and similarly, Indian YouTubers (e.g., P2, P6) also expressed disappointment about their media vilifying Pakistan. They believe partisan control over mainstream media exacerbates adversarial representations of different countries, religions, and cultural communities, leading to intolerance and communal division within countries and the region. Participant P2 articulated his disappointment about the failure of journalism—gathering, recording, verifying, and reporting information of public importance [324], on mainstream media:

*I realize that TV was dying way back in 2010-2011 because there was no growth in television. The respect in TV had gone down. I think Indian TV is talked about enough at this point in time. Even globally, you know how poison-filled and hate-filled it is. ... Pakistan bashing really gets them a lot of content and eyeballs.*

(P2, male, India)

While our participants mentioned some efforts of institutions like mainstream media to construct a diverse representation of Bengali culture and identity, they critique partisan, communal, and nationalist control by elites like politicians and media personalities, which perpetuates colonial stereotypes and communal division and outweighs its potential to contribute to the decolonial discourse. Despite the inherent politics of online platforms [231], which is out of the scope of this chapter and will be discussed in the following chapter, our participants view these platforms as alternative media to build resistance against mainstream media's hegemonic, prejudiced, and adversarial representation. They described using platforms like YouTube as a way to democratize media representation. Instead of seeing people in neighboring countries as adversaries or caricatures through a tainted political and ideological lens of mainstream media, they seek to promote humanizing perspectives to others, for which they strategize video-making on YouTube in ways described in the following section.

### **7.3.2.2 Decolonial Strategies and Reforming “Outside” Domain of Institutionalization**

Given how institutions are heavily focused on creating and propagating fragmented nationalist narratives as shaped by colonial histories, our informants participated in developing strategies to reform these institutional accounts. In doing so, their decolonial strategies target issues with textbooks and mainstream news production to offer alternative ways for people to discuss complex information and push back against the fragmentary agendas of institutionalized nationalism. These strategies include the creation of (1) political explainers on historical and contemporary geopolitics

and (2) journalistic videos about socioeconomic issues at grassroots, national, and subcontinental scales.

**Political Explainers to Make Decolonial Discourse Accessible** While political discussions on mainstream media are usually not welcoming to general audiences, our participants often structure their videos as political explainers—that explain how something works in a simple and engaging way. As we discussed in the previous subsection, regional geopolitics is one of the central factors of their transnational decolonial discourse. For the decolonial reimagination of sociopolitical structures and institutions, our participants described the accessibility of the general public to political discussions as an imperative factor. To make political discussions accessible, a few of our participants (P2, P4, P7, P8, P11) view these explainers to be an effective means. Participant P2 described his strategies to explain recent political events to his audience:

*We need to make political conversations easier. I think political caricatures really make politics fun, and I think we really need to make politics fun. ... We do political explainers. We do political satire and some humor skits—basically simplifying the news, making it more interesting for the next generation. ... I still remember the massive farm laws protests<sup>8</sup> . ... We used Lego blocks to do to explain the farmer laws. I think that people understood and remembered. So, I think there is a need to “de-serious-ize” or de-complicate political matters.* (P2, male, India)

In their approach to providing more simplified explanations of politics, our participants often focus on controversial topics. They believe that this change in choosing sources of news consumption is driven by the decline in trust in mainstream media like television and the preferences of their primary audience, which primarily consists of young people aged 18 to 35 who seek news and information presented in a conversational manner.

---

<sup>8</sup> The Indian Farm Bills, aimed to deregulate agriculture and enable direct negotiations between corporations and farmers, sparked widespread protests and was later repealed. [394].

Our participants highlight the historical and contemporary statuses of geopolitical relationships. For example, while Bangladeshi YouTubers speak highly of India's role during the liberation war of 1971, they also highlight current points of contention (e.g., The Teesta and the Ganges water dispute<sup>9</sup> , killings at border<sup>10</sup> ) in Bangladesh-India diplomatic relationship. In their videos, besides explaining complex diplomatic issues, they imagine better regional geopolitical relationships. Participant P7 underscored the importance of an improved India-Bangladesh relationship,

*[Attempt to improve Bangladesh's relationship with India] is necessary for our diplomacy, and we cannot avoid it when a large country is on three sides of a smaller country. If you have a neighbor on three sides of your home, will you want to get in conflict with that neighbor? Never. Rather, you would want to maintain a good relationship with them through exchange and equal opportunity. Also, we are India's closest neighbor. So, India would want to be on good terms with us for various reasons.*

(P7, male, Bangladesh)

Our participants often described political explainers like theirs on YouTube as productive media consumption and better use of this technology. A few of them conveyed this opinion strongly in their videos and motivated their audiences to watch more content of that sort.

**YouTube Journalism to Foreground Subaltern Experiences and Perspectives: Challenging Institutionalized Media** Some participants described their journalistic endeavors of reporting local incidents and analyzing current affairs as “YouTube journalism.” They particularly focus on localized social and political issues and practices that are often overlooked by mainstream media. They report corruption at local institutions, creating a culture of transparency at the grass-roots level. What differentiates our participants' practices from citizen journalism—the collection, dissemination, and analysis of news and information by the general public, especially by means of

---

<sup>9</sup> Teesta and Ganges are two transboundary rivers with India and Bangladesh in upstream and downstream, respectively [2].

<sup>10</sup> The Indian Border Security Force (BSF)'s shoot-to-kill policy at the Bangladesh-India border resulted in the deaths of almost 1,000 people, mostly Bangladeshis, without any prosecutions for these acts of violence [478].

websites, blogs, and social media [83], is their exclusive focus on the subaltern, through which they challenge the hegemonic political narratives and help mass people to find their voices in political and economic matters.

A couple of our participants made videos about the rural social stigma and caste-like hierarchy around certain communities and professions that impede the economic progress of those communities. For example, in one video, participant P11 interviewed female garment factory workers. While the issues these workers face are covered by mainstream media periodically after major incidents, she focused on their requirements for safety and rights for fair wages and emphasized how ensuring basic employment benefits for these workers would contribute to the national economy.

Besides identifying challenges in industrial and agricultural sectors, they also identify potentials and propose ways to include underserved communities in the mainstream of economic reformation. For example, Participant P6, an Indian female YouTuber, used the examples of mass mobile app development during the Indian ban on Chinese apps as an example of achieving technological self-sufficiency that leads to unique practices and skills of technicians with no or little educational background in the Global South around repairing devices [281, 287]. Participant P4 calls for attention to these often overlooked local industrial and postcolonial technological potentials:

*If we can utilize those technicians from Jinjira (a suburb in Dhaka) in a productive manner, we can design technology on our own terms and according to our own needs. We won't have to depend on others. If those people can develop such complex devices, albeit duplicates, without formal training, why can't they develop similar devices using their own designs? We have to support them financially and legally.*

(P4, male, Bangladesh)

Here, the participant evaluated the local light engineering talents as indicators of the country's industrial preparedness and possible hubs for sustainable technology and entrepreneurship, which researchers described as crucial for moving toward decolonial economic reformation [339, 354].

Toward decolonizing social, political, and economic structures, our participants highlight transferable lessons for Bangladesh, India, and Pakistan due to their similar sociocultural contexts. For example, Pakistani participants emphasized seeking guidance from India and Bangladesh regarding economic development, embracing their approach, and learning from their strategies for economic growth. Similarly, two of our Bangladeshi YouTubers advocated for mass digitization of grassroots economic transactions following India's mass adoption of QR codes for small businesses. In addition to economically underserved communities, Participant P2, an Indian male YouTuber, discussed subaltern communities, like religious minorities, whose protection has been a crucial issue in the region. YouTubers strategically talk about this sensitive issue, where instead of directly pointing out the bad actors (e.g., naming religious fundamentalist entities or individuals), they talk about the systematic problems in similar contexts. Participant P2 explained his approach:

*I made an episode ..., where I talked about the extremism in Bangladesh and how Hindus were being persecuted. The subtle message was that I was basically talking about [persecution of Muslims in] India while I was talking about [violence against Hindus in] Bangladesh. Basically, I was talking about minority persecution. (P2, male, India)*

In this case, the participant in his video talked about what led to rising religious extremism in Bangladesh and what steps from the government helped the minorities in crisis. However, since there is limited or politically charged coverage of these kinds of internal and external affairs on mainstream media, our participants use their YouTube channels as platforms for identifying these transferable lessons to dream, commit, and plan actions geared toward better governance.

## 7.4 Discussion

The double bind of nationalism is evident in its historical contribution to anticolonial movements while simultaneously continuing colonial legacies of division by isolating and fragmenting

regional and local identities. This chapter focuses on the relationship between nationalism and colonialism and describes YouTubers' motivations and strategies to engage in video-mediated decolonial discourse in transnational contexts. We reported how, through the construction of videos, YouTubers are engaged in decolonial practices that work to revise existing collective understandings and experiences with and of cultural and institutionalized nationalism. Inspired by Paul Dourish's seminal work [178], where he criticized the idea of technological implications as the objective of social computing research, our work seeks to highlight the participants' creative processes by which they put YouTube into practice and how this technology becomes a site for social and cultural production and provides occasions for enacting cultural and social meaning.

In this discussion, we build on these findings by discussing how technological practices get modularized as driven by sociocultural logic. Building on this, we organize further discussion, first centered around how YouTube's video-modality supports decolonial discourse. Then, we reflect on how these decolonial discourses play out beyond the online sphere.

#### **7.4.1 Modularization and Polarization in Online Discourse**

Our study leads toward understanding how creative practices around technology embed and perpetuate the understandings and experiences of nationalism of those who use them, which we dub the modularization of identities. In the context of computing, this modularity gives rise to polarization on social media—a phenomenon wherein people are engaged in political discourse only with others who are like them, diverging their political attitudes to ideological extremes [14, 515]. In the context of cultural discourse, prior research on representing regional cultural identity using videos described online platforms as the online “public sphere” [388, 389, 390]. CSCW and CHI scholarship often build on Habermas' conceptualization of the public sphere to study political deliberation and mitigating polarization [408, 505]. These works, predominantly conducted in Western contexts, described the population they studied as mostly “homogeneous” [391, 505]. While homogeneity aligns with Anderson's “modular” conceptualization of nations in the Western context, Chatterjee argued that the subcontinental imagination of nations is based on the idea of

difference [111]. The stronger the sense of differences within the broader public is, the more likely counterpublics—social groups that develop alternative interpretations of their social identities [214], are to emerge.

To complement existing work on online cultural and political expression, our study presents an interesting case from the Global South, where people strongly perceive themselves as members of several intersecting imagined communities, creating a heterogeneous platform identity. We found in our study that YouTubers making videos about Bengali identity and culture subscribe to alternate imagination of their native selves through the lenses of linguistic, religious, and postcolonial states-based nationalism. How these national and broader sociocultural identities are often defined by the exclusion of each other in the local context exacerbates the possibility of counterpublics. How do sociotechnical systems' designs respond to these intersecting imaginations of communities? What does it mean for the decolonial discourse on online platforms?

We know from postcolonial computing scholars that sociotechnical systems are predominately Western-designed products that often strive to support the development of homogeneous communities [180, 282]. These technologies, which modularize people's identities based on culturally sanctioned understandings, encounter complications as they travel and migrate to different cultural contexts. For example, after operating for almost a decade among English-speaking users, the popular Q&A site Quora expanded to other language-based platforms, with interfaces in local languages but design scaffold not equipped to handle the local cultural nuances. While prior work on Quora exploring identity decolonization found users discussing possibilities of cultural, political, and economic cooperation [152], the creation of parallel discursive areas, dubbed “stages”, to address communal and national tension bars participatory and inclusive discourse from developing further on the platform [149]. In other words, whereas local communities bludgeoned across linguistic, religious, and postcolonial states-based nationalism limit the decolonial discourse at vaguely identifying the objectives, they fall short in defining how those objectives could be achieved.

While many social computing studies champion the promise of free participation in online communities, they often overlook or downplay the challenges of intersectional tensions. Our

work foregrounds how communications across intersecting national identities or imagined communities (e.g., Pakistani or Indian YouTubers mutually exchanging views, opinions, and ideas with Bangladeshi audiences) within the broader native ethnic identity affect their articulation of social, geopolitical, cultural, and economic objectives for decolonization.

#### **7.4.2 YouTube’s Video Modality for Inclusive Decolonial Discourse**

In building on the previous subsection, our work highlighted how YouTubers were engaged in practices that served to perpetuate and support inclusive and participatory discourse. To understand how YouTube, as a discourse platform, supports this diverse exchange of opinions and ideas, we explore the features that support these discourses: (1) length of videos and (2) multimodality.

##### **7.4.2.1 Length of Videos for Creating Depth of Perspective**

Exploring similarities and diversities in Bengali cultural practices across religious and post-colonial boundaries is a major driving force behind our participants’ decolonial discourse on YouTube. Though prior works have studied how users represent and archive their cultural practices through short videos on platforms like Douyin and Kuaishou [116, 354], YouTube’s norm centered around making longer videos provides room for the YouTubers to explain their perspectives and unpack nuances and diversities within Bengali culture. For example, our participants make reaction videos where they watch a cultural artifact with their audiences and discuss different aspects of local culture based on what that artifact highlighted. Such informal discussions, in the form of a friendly watch party, resembling “*addas*” [107] function as culturally and politically important platforms for sharing ideas and experiences. Through the discussions on cultural artifacts and documentaries about history, YouTubers emphatically understand and assume others’ perspectives—what Habermas dubbed as “ideal role taking” [249]. For example, the process of a Pakistani YouTuber making a reaction video after watching a documentary on the Bengali language movement helps that YouTuber to empathize with Bangladeshis, while the Bangladeshi audiences understand an ordinary Pakistani’s perspective on their shared history. Compared to short-video sharing plat-

forms, YouTube's norm of longer videos facilitates and encourages such unpacking. Thus, our participants recognize the pluriversality–interconnection among regional experiences and views on historical milestones toward a decolonial imagination.

#### **7.4.2.2 Multimodality in Creating Accessibility**

YouTube is primarily a video-based platform. Because videos can combine modes like audio, visual, linguistic, and gestures, they can be categorized as multimodal. Moreover, YouTube supplements videos with textual elements (e.g., titles and descriptions) and spatial-visual elements (e.g., graphical user interface) besides other modes of interaction—making it a multimodal platform. Though digital content of different modalities, such as textual articles and audio podcasts, can be effective in political conversations and media activism online, the multimodality of YouTube allows for more engaging and captivating content. YouTube videos leverage the power of body language, tone of voice, and visuals to enable content creators to communicate with their audiences and effectively convey their message, surpassing the capabilities of other modes of communication. Decolonial discourse through textual communications [149, 152, 175] can be inaccessible to people who face difficulties with reading and writing in a particular language or, in general. In the post-colonial Indian subcontinent, where literacy rates remain low, YouTube's audio-visual content has the potential to reach broader audiences than textual discussions. Besides, the platform's multimodality can help viewers comprehend YouTubers speaking different languages, particularly when their written text may use different alphabets, but the spoken languages are mutually understandable (e.g., Hindi and Urdu). Thus, YouTube welcomes more diverse people and their perspectives in decolonial discourse.

## Chapter 8

### Postcolonial Politics of Content Creation

In the previous chapter, we studied how YouTubers engage in video-mediated decolonial discourse. Given colonization's deep-rooted impacts on Bengali people and their social structures, how do the audiences respond to YouTubers' such efforts? What are YouTube's roles in this process?

To answer these questions, we further examined the relationship between identity decolonization and sociotechnical systems, in this case, YouTube. We explore content creation on YouTube through the lens of postcolonial politics. Building on Winner's conceptualization of power [607], we define postcolonial politics as the distribution of power, authority, and privilege in a community historically shaped by colonization. We examined it in terms of postcolonial identities, relationships, and structures.

We explore how the postcolonial relationships among different Bengali identities impede decolonial discourse by translating tensions like communal distrust, nationalist fragmentation, and marginalization of certain identities and cultures onto digital platforms like YouTube. Our analysis also shows how online platforms conforming to economic and political power structures in postcolonial societies discourage YouTubers' efforts to engage in decolonial discourse. Our participants strategize their interaction and content creation approach to navigate these relational, representational, and structural challenges. While most HCI research that adopts a postcolonial computing

lens often tends to either overlook users' agency or glorify it [329], our research in this chapter presents the duality of postcolonial politics in YouTubers' video-mediated decolonial discourse.

## 8.1 Background

In this section, we will discuss how the content creators on YouTube are subject to platform capitalism. We will also describe how copyright policy and the multi-channel-network mechanisms their experience and video-making.

Many contemporary digital technologies, including YouTube, and their operational models are characterized by platform capitalism [539]. As a platform, YouTube connects content creators, viewers, and advertisers, extracting value primarily through data collection, targeted advertising, and service fees. Often, these data are collected through constant and comprehensive surveillance of the users [623]. While the platform maximizes its profits by personalizing user experiences and advertisements, the monetary benefits also motivate many content creators to view their success in economic terms, leading them to strategize their content [359].

In addition to the concerns regarding monetization, creativity remains an integral and inherent part of content creators' work [519], which makes content creators very protective of their content. Therefore, various provisions of the Digital Millennium Copyright Act (DMCA), such as anti-circumvention and notice-and-takedown systems [454], which protect the intellectual property rights of content creators, play a crucial role in ensuring the content creators' exclusivity to their content on digital platforms. While often viewed as a legal framework, prior works in CHI and CSCW have looked at how everyday content creators seek to understand and navigate these policies [204, 206]. Like many other digital platforms, YouTube allows copyright holders to block, monetize, or track their material within user-uploaded videos. However, users often find ways to circumvent copyrights and legal constraints in their use of technology [203]. However, this becomes increasingly difficult with YouTube's censorship policies' utilization of automated copyright detection systems (e.g., YouTube's Content ID [439]). In this process, any video uploaded to YouTube

gets scanned against an audio-video database, and a match would lead to a content ID claim. While there are ways to address and dispute these claims, before they are settled, the associated videos may be blocked. Especially for YouTubers who make transformative or remixed content that makes use of copyrighted material [208], similar to the ones making reaction or commentary videos on cultural artifacts for decolonial discourse (as discussed in Chapter 7), such deletion and blocks are challenging and discouraging. Moreover, online platforms adopt various governance practices enforced through censorship to ensure a platform's acceptability to both advertisers and users. We discussed the governance of online platforms in detail in Chapter 6. Whereas we focused on online platforms like Quora in that chapter, in this section, we focus on the different governance mechanisms in the case of YouTube, such as multi-channel networks.

Mutli-channel-networks (MCNs) are the back-end to third-party intermediaries [350]. This mechanism is not unique to YouTube, but local video-based platforms like Douyin and Kuaishou also use a similar approach [169]. However, this complex network of YouTube and content creators is integral to how YouTube operates. These are organizations that partner with multiple YouTube channels, providing support, resources, and services such as audience development, content production, monetization, and rights management [350]. MCNs assist content creators in navigating the complexities of YouTube's platform, including copyright management. They handle copyright claims and disputes, leveraging their relationships with YouTube and copyright holders to resolve issues efficiently. Additionally, MCNs negotiate better ad rates and brand deals, thereby increasing revenue for the channels they manage [242]. Through their mediation, MCNs establish a technological hegemony by gatekeeping how YouTube should operate as a platform that excludes the voices of small-scale content creators who cannot afford to join MCNs [517]. As MCNs serve as the digital intermediary of cultural logic and the mediator of legal frameworks like copyrights, this tripartite relationship shapes the politics and distribution of power on YouTube [242, 350, 517].

## 8.2 Methodological Details

### 8.2.1 Recruitment and Interviews

In this chapter, we draw on the qualitative data collected for the study presented in the previous chapter. We conducted semi-structured interviews with YouTubers from Bangladesh, India, and Pakistan. For details about how we identified, recruited, and interviewed the participants and their demographic information, please refer to section 7.2.

### 8.2.2 Data Analysis

Similar to the study in the previous chapter, we analyzed our data in an inductive, grounded theory [547] inspired approach using qualitative data analysis software MaxQDA Plus to code the interview transcripts. As per Strauss and Corbin's guidelines [547], our data had three phases. First, we iteratively reviewed the interview transcripts and identified the In the open codes—repeatedly appearing abstract concepts, events, and interactions (e.g., “*copyright claims and selective region-based blocks*”, “*uncertainty around copyright claims*”, and “*settling copyright claims*”). Then, we combined the open codes to create higher conceptual themes or axial codes (e.g., “*issues with copyright policies*”). Finally, we connected the axial codes to come up with the selective codes (e.g., “*technologically mediated power hierarchies*”) presented in this chapter. Given the contextual nature of interview data, we did not report an inter-rater reliability score [374].

## 8.3 Results

We found that various prolonged effects of colonization on Bengali people's identities and relationships and the postcolonial structures of these societies affect YouTubers' video-making and their interaction with audiences. In this section, we will discuss the challenges our participants face in engaging in video-mediated decolonial discourse on YouTube and how they find ways to continue their work.

### 8.3.1 Postcolonial Tensions in Identities and Relationships

Our participants talked about the challenges stemming from their identities that affect their relationships with the audiences and their strategies for audience management and interaction. Here, we look at how YouTubers experience and interact with their audience through the lens of postcolonial tension, which refers to the conflict and struggles for identity, power, and cultural recognition that arise in societies historically shaped by colonial rule and values.

#### 8.3.1.1 Challenges for Pluralist and Intersectional Representation

First, we will discuss various ways our participants experience challenges in building relationships and interacting with their audience. Tensions among various identities across different dimensions, such as gender, religion, nationality, language, and cultural practices, are a byproduct of this region's colonial past. Because of how colonization has historically shaped and fragmented Bengali societies, YouTubers face challenges in representing the pluralism and intersectionality of Bengali culture.

**Distrust among Religions** Previously in Chapters 5, 6, and 7, we discussed how religion has been used as the primary axis of colonial divide-and-rule policy [152], fragmenting the perception of collective identities among the Bengali people [146] and shaping a heightened sense of distrust in different religion-based communities [149]. Similar distrust and fragmentation mediate YouTubers' interaction with the audience.

For example, P1 is a Muslim YouTuber from India who makes videos about the positive aspects of Bangladeshi societies and their recent development. However, some of his audiences have alleged that his videos are promoting a Muslim-majority country (i.e., Bangladesh) because of his religious identity. How colonization has shaped the religious communities' relations in Bengal leads to such allegations among the audiences, where instead of appreciating the YouTubers' efforts, they view his videos' topics and his motivations through a lens tainted with postcolonial tension among religions. Similarly, some of the audiences of Pakistani participants P3, P10, and P13 abused them

for focusing on improving relationships with Bangladesh and India. Their audiences did not like the fact that these YouTubers, being from Muslim-majority Pakistan, wanted to improve relationships with Hindu-majority India. Despite being a Muslim-majority country, the sociopolitical fabric of Bangladesh is heavily influenced by linguistic nationalism, as we discussed in the previous chapter. Given this lack of focus on religious nationalism, while fewer in number, some audiences do not appreciate the YouTubers' effort to promote the possibility of a harmonious relationship with Bangladesh.

**Linguistic and Nationalistic Identities** While the nationalism among Bangladeshi Bengalis is primarily defined through their language, for the Indian Bengalis, Indian nationalism is the prevalent axis of their collective identity—Chapter 7 discussed this fragmentation in detail. However, given Bengali nationalism's significant and historic role in shaping Indian Bengalis' political perception, with the rise of a Hindi hegemonic nationalist perspective in the contemporary Indian political landscape, Indian Bengali YouTubers' videos about critiquing Indian politics are often questioned. Participant P2, a Bengali Hindu YouTuber from India, creates videos about contemporary politics in India as part of his desire to engage in decolonial discourse. He described how his linguistic identity is questioned and deemed insufficient to be in Indian interest.

*Bengalis are not Indian enough, so I am not Indian enough. Oh, you are [common Bengali Hindu surname]! So, there are two qualifiers [religion and language]. To them, Bengali equals Bangladeshi. ... So, I am not sent to Pakistan; by the way, I am sent to Bangladesh. So, not enough identity is something that I completely resonate with. It is all about identity—"my identity is more than your identity."(P2, male, India)*

Unlike the cases we discussed before, here P2's videos are not questioned through the lens of religion because he is a Hindu in Hindu-majority India, instead his Bengali identity is weaponized to abuse him and suspect his videos' motivation. The fragments and frictions of collective identity pose increasingly difficult challenges due to the transnational characters of most of our participants'

audiences. Generally, most of their audiences are from Bangladesh, followed by India and Pakistan as the second and third largest bases of their viewers and subscribers. They also had significant numbers of viewers from the Gulf countries (e.g., Saudi Arabia, United Arab Emirates, Kuwait, Oman), the UK, and the USA, which they believe are Bengali diaspora. When our participants' discourses deal with broader regional issues encompassing both the positive and negative roles of neighboring countries in geopolitical relationships, sometimes, a part of their audience questions our participants' knowledge on those matters out of confirmation bias and chauvinist attitude. For example, P9 is a history-enthusiast Indian YouTuber who makes videos about precolonial, colonial, and postcolonial history of Bengal. Since educational systems in different countries present fragmented historical narratives (as we found in Chapter 7), people often internalize a particular historical narrative. Hence, many of P9's Bangladeshi viewers who learned the dominant narrative about a certain individual or event in Bangladesh suspect and question the information he presents in his video. P9 described his experience:

*They would say that I was spreading misinformation. Then, I asked them to give the correct information. Then, they could not respond. There are also times when something is in history, but they are not willing to accept it. They hold onto their personal biased belief, and they won't come out of it. ... I aggregate historical information to make a script but do not put my own opinion there. ... But some people are strongly opinionated and won't come out of their beliefs. Whatever does not align with their existing beliefs, they are not willing to accept that. (P9, male, India)*

Again, the tense relationships among Bangladesh, India, and Pakistan often made YouTubers afraid to make videos on geopolitical issues. While most of our participants gathered enough courage to make such videos regularly, Pakistani participant P10 shared that some Bangladeshi YouTubers who were scared contacted him. He encouraged and advised them to make videos while being mindful of political safety. While in the previous chapter, we found a lack of interest in political

discussions, our participants (e.g., P13) also talked about the direct influence of politics on their channels' subscribers:

*When the standoff with India happened, there was a campaign of Indians unsubscribing Pakistani channels, and we were victims of that as well. We lost our 4000 subscribers at that moment.* (P13, male, Pakistan)

This is an example of brigading on YouTube. Here, the audiences band together based on their postcolonial nationalist identities and lead coordinated activity to deliberately target YouTube channels run by content creators who are not part of their in-groups. Similar to our discussion in Chapter 7, here, the out-group and in-group identities among Bangladesh, India, and Pakistan are defined by their histories of postcolonial partition and independence. Such brigading on online platforms (e.g., Reddit [175]), groups of users disrupt specific posts, users, or communities to influence the conversation, manipulate voting systems, or harass individuals. In this case, such behavior fueled by postcolonial nationalism disrupts the normal flow of discourse and can skew the perceived popularity or reception of content focused on the decolonization of sociopolitical relationships in the region.

**Subaltern Gender and Culture** Postcolonial scholars interrogated the representation of people from different intersections of colonially marginalized societies in colonial societies. For example, subaltern women's voices were often silenced or misinterpreted by dominant discourses by both colonial and Indigenous patriarchal structures [537]. We found that similar postcolonial politics influence our participants' experience and interaction on YouTube.

Similar to the Bangladeshi Bengali accent/dialect being harassed because of its similarity with the refugees who migrated to India post-partition (as we discussed in Chapter 6), our participants often faced ad hominem attacks for their facial appearance, accent, and dress. As colonization shaped the ideas of masculine Englishman and effeminate Bengali man [524], it exacerbated patriarchy in Bengal by bolstering traditional gender norms and marginalized women. For example, in Bangladesh, the Muslim-majority nation-state emerging from the postcolonial partition, expecta-

tions about women's ways of being (e.g., dress, gesture, speaking style, comments on social media) are shaped by the concept of *Purdah* in Islamic culture [460]. We found that the audiences have particular expectations about how female YouTubers dress. P15 talked about several instances of her Bangladeshi male audiences categorizing her sleeveless dress as short and abusing her with vulgar comments. All four of our female participants talked about the gendered aspect of the harassment, obstruction, and fear they face in their content creation. Their hurdles are higher for making videos on issues like grassroots political and economic corruption, which is an important part of the online decolonial discourse [146]. Participant P11, a Bangladeshi female YouTuber, shared that a report she wanted to post from her channel required video shooting at night. However, due to the greater risk to her safety as a woman, she could not go. She said:

*Many people would certainly obstruct me. They will scold me, saying, "You are a woman, and you came to make videos!" They behave the way they were shaped to behave in Bangladeshi culture. ... I cannot go to collect footage for those videos at night.* (P11, female, Bangladesh)

In the previous chapter, we discussed how the standard form of linguistic practices and culture in colonized and postcolonial Bengal was strongly influenced by native Bengali elites while putting the people in the agrarian East Bengal in the unvoiced subaltern space [111, 112]. Our participants also echoed prior scholarship on the socioeconomic stratification of culture in postcolonial settings. For example, P14, a participant from India who makes videos about Bengali culture in Bangladesh, said:

*Some rebuked us for highlighting [a popular Bangladeshi media personality]'s work. They told us that people in cities do not follow her work and that they are only for the uncultured village people. They used exactly such language.* (P14, male, India)

Here, some viewers of P14's videos have a presumed notion of what Bengali culture entails. Historically, in colonial Bengal, urban *Bhodrolok* (civilized) practices were normalized as the stan-

standard form of Bengali culture while marginalizing the rural people as *Chhotolok* (uncivilized)[112]. In P14's quote above, we can see the viewers' colonially influenced understanding of the "standard" form of Bengali culture creates challenges for decolonial discourse through reactions or commentary videos on cultural artifacts.

### 8.3.1.2 Strategies for Resilience and Audience Management

In this section, we will discuss how our participants continue making videos for decolonial discourse on YouTube. They often adapt to the challenges by finding support from their social networks and audiences, following the identity norms and customs of their targeted audience, and deciding on the dilemma of content topics.

**Finding Support from Community** While our participants find it challenging to continue making videos in the decolonial discourse space for various reasons, they also appreciate the support they receive from their families, friends, and audiences. This is a discourse strategy for a person to find resilience through the comfort and support of the community. Indian participant P12 described her audience's cordiality:

*Some say, "Sister, if you come to Bangladesh, come and stay at our home. We will take you to visit different places in Bangladesh." It seems like they are my own family members. When I read their responses, I get such a feeling that one can expect while interacting with one's siblings. This encourages me a lot.* (P12, female, India)

In addition to the encouragement online, given the gravity of the female YouTubers' lack of safety in real life, trusted people in their existing social network and the ones who appreciated their content online often offer help. For example, in P11's case, as described above, when she needed to shoot videos at night but as a woman did not feel safe, some of her male friends came forward to help her. Our participants sometimes recognized their volunteer members with nominal financial compensation. These volunteers' support for our participants in developing and sustaining

videos about sensitive topics can range from accompanying YouTubers in large groups to arranging special safety measures (e.g., dedicated police protection). Participant P11 described one of her past experiences as follows:

*We consider [...] if we can go there safely. There are many alumni from our university in high levels of administration. They can be high-ranked bureaucrats or police officers, but for us, they are like our brothers and sisters. ... We also attend to the financial needs of [the junior friends] who help us from our channel's income, although they never ask for it.* (P11, female, Bangladesh)

**Following Identity Norms and Customs** Given the fragments and distrust among different identities within the Bengali people, our participants try to connect and find commonality with their audiences through the performativity of different identities. Scholars argue that identity is constructed through repeated actions, behaviors, and performances that align with societal norms rather than being an innate or fixed trait [100]. Thus, by adopting and conforming to different religious or national communities' linguistic practices, norms, and customs, YouTubers shift the audience's focus from their birth religion or nation-states. For example, given the variation and strong association of salutations with different religious communities, P3 uses common phrases for Bengali communities and specific ones for Muslims (e.g., *Assalamualaikum*) and Hindus (e.g., *Namaskar*). Some of our participants also start the videos with the primary language or Bengali accent (e.g., *Bangal* or *Ghoti*) of their audiences. However, since some of our participants (e.g., the ones from Pakistan) do not speak Bengali as their native tongue or the accent of their audience (e.g., Indian Bengali participants who do not speak the Bangladeshi form of the language), they also try to be mindful of not mispronouncing the languages. They often provide subtitles in different languages (e.g., Bengali, English), considering the transnational aspect of their subscribers and viewers and the potential for increased monetary potential. Based on their understanding or folk theory about how YouTube works, they write the countries' names in their languages so that those

are often recommended to their audiences. Some of them also take help from their friends who speak the languages natively in this task. For example, P10 explained:

*Sometimes, I try Bengali in the beginning. It attracts the viewers and makes them feel connected to our channel. It's a kind of loving gesture from us, pronouncing your language and starting with. ... My father is concerned that the Bengalis feel proud of [their language], so I must pronounce it correctly. The [Bengali] words I learned took me 4-5 months to correct the pronunciation. ... My friends from Bangladesh put subtitles in Bengali.* (P10, male, Pakistan)

**Deciding on the Dilemma of Content Topics** To mitigate the identity-based tension, our participants establish boundaries on what topics they make videos on. They decide on this through negotiation of what topics they want to focus on and what topics their audience prefers and finds less controversial. As elaborated in this section, this affects their content's success differently in ideological and pragmatic terms.

Some participants who are passionate about the decolonial aspect of their video-making on YouTube come up with various creative strategies. Considering ideological stubbornness as one of the main impediments behind political discussions, participant P2 came up with a caricature that portrays a blind supporter of certain ideologies. Through that character, he tried to speculate how his videos could be interpreted in different political echo chambers. In his words:

*I didn't do [caricature's name: X] for the sake of doing it. I used to foolproof my tweets as a blind supporter of an ideology: how those could be taken out of context, how those could be misread, how they were going to say that "Look, what he's talking about is insulting India." ... I used to think about what logic someone like X would apply to this, so that is how the whole concept came around. Of course, my school and college acquaintances who share characteristics of that caricature in*

*the WhatsApp group gave me enough understanding of how they think.* (P2, male, India)

However, most of our participants often choose to avoid political and religious discussions. While YouTubers have greater agency in deciding the topics for making videos, they cannot completely control what topics their viewers talk about. Therefore, some participants chose moderators from their regular audiences to manage their channel-associated Facebook groups (more on this later in this section). This kind of delegation gives their audiences a sense of agency. Some participants believe this delegation of decision-making on what topics are discussed in the channels brought a change of heart in some viewers who previously were negative about their work. For example, P13 started shifting away from making videos about politics himself, whereas P14 and P15 forbade any kind of religious and political content and discussion altogether in their groups. However, prior works [146, 152] have highlighted discussions around these fragmented identities crucial to online decolonial discourse. Thus, avoiding these discussions while our participants reduce toxicity, remain appealing to a broader audience, and secure their monetary potential defeats these interactions' potential for identity decolonization.

### **8.3.2 Postcolonial Structures Mediated through Technology**

Our participants discussed various challenges that emerge from the mediation of postcolonial economic, institutional, and political structures through YouTube. We also explain how our participants overcome, overlook, and navigate through these challenges by strategizing their use and presence on online platforms.

#### **8.3.2.1 Economic, Institutional, and Political Challenges**

This section discusses the challenges YouTubers face from the platform because of its differential and non-transparent policies, complicity in establishing hierarchies among cultural elites and subalterns, and coordination of political control on the YouTubers in the Global South.

**Differential and Non-transparent Policies** YouTube's policies determine what content YouTubers can create, how their content is moderated, and how they are compensated. Our participants discussed how YouTube's monetization policy is discriminatory and provides significantly different financial advantages to them based on the location of their viewers. They explained that since more companies in the Global North advertise on YouTube than the ones in the Global South, the ad revenue generated in the Global South is lower than that in the Global North. Since one of the motivations for YouTubers is to generate earnings from their content on YouTube, the platform capitalism strongly influences the YouTubers' decisions around for whom they make content. For example, within the Bengali communities, our participants prioritize making videos that interest the Bengali diaspora more than the local Bengali communities in Bangladesh or India. P14 explained:

*If someone watches a video from the US or the UK, our earnings will be higher than that of someone watching it from Bangladesh. The ads on a video depend on the country [where one is watching from]. The earnings will depend on the prices of ads in that country.* (P14, male, India)

Since YouTubers prioritize the cultural values and practices of the Bengali diaspora living in the Global North, this normalizes the representation of a particular set of Bengali cultural preferences, practices, and identities online. The Bengali diasporas in the US or the UK include people who migrated quite a while ago [10, 603], and the ones migrating in recent years for better opportunities usually come with higher education or from privileged socioeconomic classes [276, 401]. Thus, the diasporas in these locations represent an archaic or selective set of Bengali cultural preferences, which does not portray the shifts in local Bengali culture in Bangladesh and India. This also exacerbates the marginalization of some (e.g., rural, agrarian) Bengali communities to a subaltern space. While the representation of diasporic practices is important, in prioritizing the diaspora for the sake of higher financial incentives, the YouTubers commodify their viewers from a commercial perspective, which Aimé Césaire described as one of the defining characteristics of

colonial values [102]. Consequently, the platform’s differential monetization policies shape what our participants represent in their discourse on Bengali sociocultural topics and issues.

Contrary to the rationales about how YouTube monetized the videos, our participants were often unclear about many other policies on the platform. They were often frustrated about the opaqueness of content moderation (e.g., deletion or blocks of content). For example, P14, a participant from India, who primarily makes videos about Bengali culture’s transnational aspects, i.e., how the Bengali culture in India varies from that in Bangladesh, wondered why some of his videos are blocked in Bangladesh while viewers from the rest of the world can see them. Such uncertainty around whether the audience for whom or based on whose requests the YouTubers made certain videos would be able to see poses a challenge for them. They often emailed YouTube asking for clarification about its decision on their content. However, YouTube’s response is often delayed and unhelpful. Prior works have described such non-transparent policies in content moderation and governance as part of algorithmic coloniality and a bar for marginalized communities [149, 229].

In addition to monetization and content moderation policies, our participants expressed concerns about how YouTube deals with copyrights on the platform, which we will discuss next.

**Complicity to Hierarchical Cultural Logic** YouTube’s processes and general practices around copyright claims create challenges for several of our participants who ran smaller channels. Many of them make “reaction videos” where they watch entertainment artifacts like television series episodes, film trailers, music videos, and short documentaries and show their emotional reactions while watching. In Chapter 7, we discussed how this genre of videos is popular and important for decolonial discourse in the Bengali context and an accessible starting point for new YouTubers [146]. However, because they rely on re-using parts of previous videos published by others, these videos are prone to copyright claims. Our participants shared that while some channels are lenient with such content creators, others are more aggressive in claiming copyrights. Such exception to copyrights is related to fair use, which allows one to reuse copyright-protected materials without permission from the copyright owners [131]. While there are guidelines on what counts as

fair use, it is determined on a case-by-case basis by the courts and not technological entities like YouTube [134].

Our participants seemed to not know about this, and they criticized YouTube for taking “a hands-off approach to settling copyright claims.” Given that non-infringing videos are known to be flagged sometimes by content ID [536], YouTube currently has well-outlined processes for disputing and appealing copyrights, such as the ones flagged by content ID, which looks at criteria like eligibility, reason, requirements, and rationale [135]. However, none of our participants who complained about YouTube’s non-mediation talked about these processes put in place by YouTube. This indicated their lack of understanding of YouTube’s claim dispute-and-appeal processes and the platform’s inability to determine fair use. They said that the platform suggests that content creators talk to each other with the promise to abide by their mutual decisions. At this point, our participants highlighted how institutions like multi-channel networks (MCN) impact their work. In their opinion, usually more reputed, bigger, and monetarily more solvent YouTube channels join MCNs. They described incidents where an MCN-member YouTube channel used the footage from our participants’ videos and later claimed copyrights for those videos. Participant P4 critiques YouTube’s copyright policy as “weak” and describes how the prioritization of the claims from MCN member channels discriminates against them:

*YouTube blindly trusts one who has MCN. ... YouTube believes that a channel has MCN, so whatever that channel uses belongs to [that channel]. ... You can sometimes be blamed under copyright policy for using your own content.*(P4, male, Bangladesh)

While MCN serves as the digital intermediary of cultural logic on YouTube [350], based on such incidents, our participants found YouTube to be complicit in reflecting the economic hierarchies onto Bengali culture. This aligns with postcolonial scholars’ critique of how the local elites’ cultural and historical perspectives have been prioritized in colonial and postcolonial Bengal [111], often

leading the accounts of the ones from underprivileged socioeconomic classes being to silenced or categorized as unreliable [64].

As the frustrated participants perceived a lack of responsibility from YouTube for mediating copyright claims, they viewed YouTube as a postcolonial technological complicit to the hierarchical cultural logic. Especially given the power hierarchy created among YouTube channels through MCNs, our participants propose that YouTube plays a more active role in these matters and utilizes the videos' metadata. P11 shared her concern and proposal as follows:

*Do you think the other party who stole my video will tell YouTube that they stole my content and they are the ones at fault? ... YouTube can easily check who uploaded the content first, but they do not do it. I uploaded a video first, then someone else downloaded it and uploaded it. They make me the party at fault with their power of MCN.*

(P11, female, Bangladesh)

Thus, while there exist ways to adjudicate and settle concerns around copyrights (e.g., fair use) through legal processes and YouTube provides pathways (e.g., counter notification) toward that in the appropriate cases [135], societal power reflected through MCN highlights postcolonial computing's argument about how designs, in this case, of the content creation spaces for decolonial discourse, are power-laden [282].

**Liaison of Political Regulation and Control** Decolonial scholars have explained how postcolonial nation-states' political power and institutions reanimate colonial values [111, 448] (discussed in detail in Chapter 7). In doing so, some critiqued the forms of hierarchy and authority in governments as the consequences of colonization that perpetuate colonial values [448].

Our participants explained how MCNs become the liaison of political regulation and control of the government on the YouTubers. While prior research highlighted how YouTube could serve as a digital public sphere and a space for discussing political issues more freely than in mainstream media, our participant discussed how the governments in different nation-states, Bangladesh, India,

and Pakistan, are gradually exercising regulations on YouTube content creators. To explain the mechanism of how such political control is imposed through MCN, P2 said:

*The government then realized that it was getting out of hand, so every news YouTuber technically needs to be registered with the government now, and they need to be registered with a particular agency. Something that we are affiliated with is called [an MCN's name]. It has [X number] news outlets under its wing and this umbrella body, ... which makes representations to the government, or if the government wants to send a message down to you, it will use this body. So gradually, from a free speech area, we are coming under the control and into the regulated area.*

(P2, male, India)

With political control institutionalized through YouTube and MCN, people become fearful about making videos that critique and interrogate the political corruption, injustice, and status quo. Hence, our participants who had larger and more well-known channels faced challenges finding like-minded, willing, skilled, and talented human power. This poses multi-faceted and more challenges for YouTube channels focused on an ideological objective than those exploring different video topics and genres for better monetization. Participant P2, an Indian YouTuber who makes political explainer videos, said:

*It's always very difficult to find people who would be willing to work on such issues. Ideologically, many people might not agree with you and might not want to work with you. Risk-wise, it's not the safest profession to work in many ways. It's not really raining money, so people who really believe in the stuff that we do usually work with us getting facts, getting information, getting scripts is always going to be a problem in this kind of work. I've tried to find people who could write funny scripts, which doesn't come easily. ... We can manage the editors or resources, but*

*the brainpower to run the satire channel continues to be a problem.* (P2, male, India)

While the rationale for YouTube’s accountability to local and national authorities is understandable, the scope of YouTube’s cooperation with political institutions raises concerns about reinforcing postcolonial power structures through technology.

### 8.3.2.2 Strategizing Use and Presence on the Platforms

Our participants negotiate among the demands of their audience, the potential of monetization, and the possibility of their videos being blocked or claimed for copyrights. They often strategize their video-making by prioritizing non-monetary incentives, acting based on collective folk theories, and utilizing an ecosystem of multiple platforms.

**Prioritizing Non-monetary Incentives** Because of YouTube’s monetization policy, our participants were likely to earn more by making videos appealing to viewers in the Global North. Considering the financial potential, making videos for viewers in the Global South locations is not beneficial for them. However, most of their audiences come from countries like Bangladesh and India. Therefore, though the ad revenue per watch from these countries is low, making videos that interest the viewers there often entails more engagement (e.g., like, comment, share) in their videos. Therefore, they often shift from focusing on the monetary gain from a particular video to using it as an opportunity to invest in their channels’ future growth and reputation.

Reaction videos are quite popular among our participants’ Bangladeshi and Indian audiences. Despite these videos’ lower earning potential due to the viewers’ geographic location and the possibility of larger YouTube channels or MCNs claiming copyrights for reusing their content, our participants (e.g., P1) believe that making these reaction videos demanded by their viewers help them “build affinity” with their audiences. P15 explained her rationale:

*In some cases, so many viewers request us to make a reaction video that we must make it. All our hard work behind that video gets lost because those videos are not*

*monetized. Even if we sacrifice the money, it will attract many people and help us gain more subscribers. Therefore, although we are likely to get copyright claims, we make some videos to gain subscribers.* (P15, female, India)

By conceptualizing a video's success in terms of non-monetary incentives, our participants prioritize the objective of getting more subscribers and engagement over earning more dollar amounts. Using this strategy, YouTubers overlook the monetization challenges stemming from the platform's policies.

**Acting based on Collective Folk Theories** Due to YouTube's lack of transparency in its policies, the participants discussed their experiences of penalized monetization or blocks with other content creators. Through this, they developed collective folk theories about the reasons for certain behaviors or decisions by YouTube's algorithms and policies. For example, P4 explained one of his team's past experiences. They made a video report on a local corrupt officer. To protect the privacy of a minor victim, in that video, they blurred their face. When that video was yellow monetized, i.e., limited or no ads from all advertisers, they wanted to understand its reason. He said,

*We blurred a kid's face because we did not want to show them. ... As soon as we blurred their face, YouTube Yellow monetized our video. ... [A person consulted with on this matter] told me that I have a blurred face in my video and suggested to remove the blurred face. ... We did not know YouTube would do so. It is not mentioned anywhere in their policy. Some things with YouTube's policy are eccentric – they do whatever they want.* (P4, male, Bangladesh)

Thus, our participants devised various ways to circumvent these hurdles as understood based on their collective folk theories. For example, they have found that while using other YouTube channels' content in their videos (e.g., in reaction videos), sometimes providing detailed credits to the original content creator channel is helpful to avoid a copyright claim. However, giving credit to

the copyright owners or adding disclaimers about non-infringing intent does not imply fair use, as there were many cases for our participants where giving credit to the original content creators was not enough. A common pattern of reaction videos is to display the original content YouTubers react to or comment on as a smaller video within their newly made video. To avoid automated detection of such reusing of others' content, they make the appearance of those contents semi-transparent or low-resolution. They argue that this strategy, besides giving credit to the original content creators, would encourage people to visit the original content creators' channel for better video quality and, at the same time, make their videos make sense to their viewers. They also recommended that considering the popularity of reaction videos among Bengali communities, YouTube could allow them to use previous YouTubers' content up to a certain length or fraction within their newly made videos (e.g., displaying others' content for 2-3 minutes with proper credits in a 20-minute video) without infringing others' copyrights. However, the grounds of fair use—new content being transformative enough and adding new expression and meaning to the original content, cannot be determined by YouTube. Since it is determined by local legal systems and YouTube cannot take some straightforward ways to solve the concerns around it, communicating the fair use policies and copyright claims, dispute, and appeal process in various local languages can be useful for content creators in the Global South.

**Utilizing a Combination of Multiple Platforms** Based on the knowledge of online platforms being subject to different regulatory and political institutions, a common strategy among our participants was to build a multi-platform ecosystem or combination for disseminating their videos. P2 explains his strategy to resist technologically mediated political influence through a multi-platform presence:

*[YouTube] can de-platform you at any given time without warning. I mean, today, the government can write one letter and say that YouTube shut this [channel] down and they will shut it down. So, it is never a good strategy to be one platform. Each*

*video we make on YouTube, we upload it to on Facebook the next day.* (P2, male, India)

Our participants also have a dominant perception that various online platforms comply with and enforce certain rules differently. For example, they often share videos on Facebook, which they believe “is much more easygoing in terms of copyright than YouTube.” If their videos are deleted from YouTube because of copyright claims or governmental decisions, the videos shared on Facebook help them prove to their audience that they indeed made certain often-requested videos and ensure that their efforts do not end up in vain.

Such a multi-platform approach also helps YouTubers reach viewers of diverse demographic backgrounds. As the same content regurgitates itself into multiple platforms (e.g., YouTube, Facebook), it gives YouTubers multiple revenue streams, helping them overcome the challenges of lower monetization on the Global South-related content. While some participants use multiple platforms to post identical content, some strategize using different platforms for various purposes, such as Facebook groups to build community with their audiences (e.g., P12, P14, P15), Facebook pages to share promos of upcoming YouTube videos (e.g., P1, P9, P10, P13), Twitter to test water on contemporary events and post one-line punches (e.g. P2, P10), and Instagram is to do fun stuff like memes (e.g., P2).

Some of our participants, in addition to their presence in multiple online communities, set up online financial platforms (e.g., Patreon, Paytm, and PayPal) associated with their YouTube channels. Different strategies (e.g., live sessions, merchandise) have brought better monetary support for our participants. Besides overcoming the financial challenges, these premium channels have helped them to avoid the abuses stemming from the postcolonial tensions we discussed earlier. P2 explains this dual benefit as follows:

*Live session acts as a revenue source, also because I answer questions that are paid for. What I have discovered is people don't pay you money to abuse you, so I get a good bunch of questions from people.* (P2, male, India)

## 8.4 Discussion

Our participants' decolonial discourse on YouTube is a discursive space mediated through narrative and practices. In their work, YouTubers discuss Bengali culture and its history. They make different kinds of videos, such as reactions/commentary on cultural artifacts, travel vlogs, political explainers, satire, local journalistic reports, etc., as part of their decolonial discourse. To navigate various postcolonial politics affecting their content creation, they come up with various strategies. Encompassing their YouTube channels, other online platforms, and audiences, we can characterize their ecosystem of content creation as a designed sphere.

HCI researchers have examined why designing projects for social justice fails within the capitalist paradigm of technology [171, 609]. Decolonial discourse being driven by a motivation for social justice also faces challenges and conflicts with the capitalist values of platforms like YouTube. For example, in our study, we learned about the motivation of monetary incentives among YouTubers and the significantly lower monetization from YouTube for content viewed in the Global South. Increased monetization requires YouTubers to maximize engagement from certain geographic locations. Therefore, decolonial discourse, which often involves in-depth, critical examination of historical and contemporary power structures, does not typically elicit the same immediate engagement. Our participants were aware that political conversations are more likely to be controversial or less engaging. Moreover, advertisers typically prefer content that is non-controversial and broadly appealing. They might view decolonial discourse, which often challenges established social and economic structures, as potentially divisive or contentious. Therefore, except for a few YouTubers who were really passionate about the decolonial objective, most YouTubers wander among sensational and entertainment-focused topics to better capitalize on their content.

Whereas decolonial discourse seeks to challenge and deconstruct dominant narratives, the recommendation systems on online platforms tend to shape echo chambers. In addition to the struggles to penetrate these echo chambers that limit decolonial videos' exposure to a wider audience, decolonial content frequently encounters resistance from segments of the audience who

perceive it as a threat to their worldview or social standing. Our study highlighted how the audiences' hostile responses, negative comments, and brigading create a discouraging environment for YouTubers. Again, YouTube's complicity with political institutions to reach the next billion users in the Global South will pose a higher risk for YouTubers' personal safety, especially when they challenge political malpractice, corruption, and marginalization as part of their decolonial work of interrogating postcolonial power structure. Creators are thus discouraged from engaging with decolonial themes. While our study in the previous chapter highlighted how YouTube motivates and enables decolonial discourse, this chapter critically looks at the ways in which the postcolonial constructs of Bengali societies reflected on YouTube, combined with the capitalist values embedded within it, choke the decolonial potential of such platforms.

## Chapter 9

### Coloniality of Algorithmic Tools and Datasets

Natural language processing (NLP) enables computers to “understand,” “interpret,” and “generate” language. One kind of NLP is centered around analyzing “sentiment,” which is the process of determining the emotional tone expressed in text data. Applying NLP tools designed in the West to other language and cultural traditions can undermine “safety measures” (e.g., in content moderation) [411, 412] and impose Western values and perspectives. Since artificial intelligence (AI)-based technologies disproportionately harm marginalized communities like non-native English speakers [11, 455], researchers have called for increased focus on non-English NLP studies [34, 412].

In this chapter, we employ a sociotechnical approach to exploring NLP tools and their biases. By “sociotechnical,” we are not referring to a specific tool or set of technologies/tools but highlight that technology shapes and is shaped by human action and interaction [480]. We know from prior work that artifacts like algorithms and machine learning (ML) technologies are political and are shaped by societal norms as well as the individual or developer group’s politics within which they are designed [486, 607]. Sentiment analysis tools—which seek to assign subjectivity or polarity scores (usually within standardized scales) or nominal sentiment categories (e.g., positive, negative, neutral), in particular, are sociotechnical [581] in how they shape and are shaped by human interaction. It is an exercise of quantifying and categorizing complex human language and emotion. However, researchers have highlighted how sorting and categorizing are political and reductionist and can perpetuate inequality [79, 180]. On the one hand, people develop these tools,

and user interaction data is often used to train these tools, which shapes their outputs. On the other hand, when deployed in downstream tasks, which are practical applications leveraging the outputs of these tools, such as decision-making in content moderation [553, 574], their influence on user interaction is pivotal in determining permissible user interactions.

When such processes are used by computing systems to interpret and analyze human language, their outcomes often include social and technical biases [215]. Critical algorithmic studies scholars defined bias as when computer systems consistently and unfairly discriminate against certain individuals or groups in favor of others [215]. Social power structures, global resource availability, and biases can manifest in various ways through computing systems. Especially in NLP, there is an incredible disparity in research and resources available across various languages. Joshi and colleagues identified 0.28% of languages as “the winners” and 88.38% of languages as the ones “left behind” [301]. For example, although English and Bengali have comparable numbers of speakers [335], English has hundreds of times higher visibility than Bengali in terms of resources on Linguistic Data Consortium, Wikipedia, and publication venues [301]. Besides the resource disparities across languages, attention to how bias works in non-English systems has not been explored. Imposing insights about bias in Euro-centric (e.g., English) language technologies on diverse user communities without considering their cultural and historical contexts can have deleterious impacts.

As people continue to adopt computational linguistic systems, the possibility of propagating harmful decisions made with their assistance can have downstream effects—consequences experienced at later stages. Therefore, it is incredibly important to understand the application of NLP in non-Western settings. To address these myriad concerns, our research foregrounds non-English NLP research, particularly sentiment analysis in Bengali, from the perspective of fairness and bias. We investigated how Bengali sentiment analysis (BSA) tools and fine-tuned language models with BSA datasets assess specific identities, explore differences in their responses for explicit and implicit identity expressions, and examine potential biases across different identity categories and the relationship between bias and tool developer demographics.

We conducted an algorithmic audit of BSA tools available on Python Package Index (PyPI) and GitHub, BSA datasets found on Google Dataset Search, and language models (e.g., with different algorithms/architectures and training data with different breadth and diversity). Looking at different genders, religions, and nationalities, we found that different BSA tools and fine-tuned models assign significantly different sentiment scores for identical sentences expressing a particular identity. In particular, BSA tools often rate an explicit expression of Bengali identity based on nationality more negatively than when the same identities’ implicit expression through linguistic norms and regional dialects. We also found the majority of tools to be biased. Among the 13 tools we audited, 38% and 30% are respectively biased toward female and male gender identities, 30% and 38% are biased across religious (e.g., Hindu and Muslim), and 77% and 15% were biased across nationality-based identities (e.g., Bangladeshis and Indians)—reanimating the colonial hierarchies. We also found that among 38 combinations of fine-tuned models based on two pre-trained language models and 19 BSA datasets, 24% and 61% exhibited biases toward female and male identities, 24% and 61% to Hindus and Muslims, and 47% and 26% toward Bangladeshis and Indians. Though we found a digital divide among diverse Bengali communities in developing language technologies, our analysis did not suggest that the demographics of the tools’ or datasets’ developers conclusively affect the bias within sentiment tools. Taken together, our work highlights how BSA tools, datasets, and algorithms exhibit a “colonial impulse.” We discuss the downstream implications of using available BSA tools and datasets in content moderation and fine-tuning models for specific tasks.

## 9.1 Background

In this section, we will discuss how various social identities are expressed through language and technologies. As technologically represented identities become subject to algorithmic systems, the latter part of this section will also describe how algorithmic audits deconstruct these biases.

### 9.1.1 Expressions of Social Identity through Language and Technology

Coloniality has continued to shape people's everyday experiences and, on a deeper level, mediate how they express their social identities. One can express one's social identity both explicitly and implicitly. Explicit expressions of identity refer to deliberate and direct ways individuals communicate and assert their affiliations, characteristics, and beliefs. For example, mentioning one's nationality and political views or openly discussing one's religious beliefs are examples of explicit expressions of identity [560]. Meanwhile, implicit expressions of identity include subtle and indirect ways in which identity is communicated or inferred from a person's actions, behaviors, choices, and interactions [568] and are bound up with cultural norms, societal expectations, and institutionalized practices [100, 270]. For example, how one speaks, the words they use, or their hobbies can implicitly give insights about one's identity. While people's social identities can be communicated implicitly through different speech and non-verbal acts, this chapter focuses on linguistic expressions of various identity categories through writing. Particularly, we considered how different gender, religion, and nationality-based identities are expressed explicitly and implicitly in Bengali texts.

Cultural-linguistic scholars have detailed how languages are often standardized differently in different countries (e.g., English in England vs. the United States; German in Germany vs. Austria) [88]. These geo-cultural variations, often referred to as dialects, operate as important signs and implicit expressions of cultural identity [196, 263]. In Bengali, the two main dialects are *Bangal* and *Ghoti*, which are spoken in East Bengal (Bangladesh) and West Bengal (in India), respectively [147]. These variations of the Bengali language manifest both phonologically and textually [316, 432] and use different colloquial vocabularies in written texts for the same everyday objects. For example, Bangladeshi and Indian Bengalis respectively use the words "জল" (/ʒɔl/) and "পানি" (/pa:ni/) to mean "water." Consistently using vocabulary from either the Bangal or Ghoti dialects can implicitly express a Bengali person's national identity without any explicit mention. Similarly, Bengali textual communication often implies the gender and religious identities of the

people it describes. While in Bengali, unlike many other Indo-European languages, gender does not change the choice of pronouns (as in English) and verbs (as in Hindi and Urdu) [66], culturally, most names and kinship terms are gender-specific with some exceptions [167]. Moreover, commonly used kinship terms, names, and commonly used vocabularies often implicitly indicate one's membership or being born into either Hindu or Muslim communities [147, 167]. For example, while Bengali Hindus often draw inspiration from Demigods' names and characters in legends for their personal names and commonly tend to use Bengali words derived from Sanskrit, in Bengali Muslim communities being named after Prophets, Caliphs, and Mughal emperors and the vernacular use of Perso-Arabic words are widely popular [167]. Thus, written Bengali communication can lead to the inference of one's gender, religion, and nationality-based identities.

As the colonizers invented categorization and classifications by viewing and interpreting cultures, societies, and people from non-Western locations in a stereotyped and exoticized manner [473], hierarchies among these artificial categories have been established and embedded within colonized societies [152, 199]. Broadly, these experiences included everything from colonially shaped racism (a belief in certain racial groups' inherent superiority or inferiority) to colorism (favoring lighter skin tones over darker ones within a single racial group). With respect to how people express their social identities through written language, the influence and affluence of West Bengal's upper-caste Hindu landlords and elites, who predominantly spoke the *Ghoti* dialect, led to the establishment of their dialect as the institutional and "normative" standard for the Bengali language during the introduction of printing presses in the region [111]. In contrast, the *Bangal* dialect became associated with East Bengal's agrarian socioeconomic system and refugees due to mass migrations following the colonial partition and a means of Muslim and underprivileged caste Hindus' social harassment [149, 225]. Through coloniality, these impacts on identity, such as sociolects (dialects of particular social classes [373]) and colonial ontologies and epistemologies—the ways of being and knowing—are embedded within the world structures at regional and global scales and continued across generations through various artifacts, media, and technology [17, 47].

This leads to critical and important questions: Are sociotechnical systems “mindful” of such sociocultural and historical complexities that shape people’s identities? How are identities translated into “something a microchip can understand” [470]?

### 9.1.2 Algorithmic Bias Deconstruction in Computing Systems

To better interrogate these questions, we draw on postcolonial computing scholarship. Broadly construed, postcolonial and decolonial scholars have worked to highlight the “colonial impulse” of technology [180, 282]. Dourish and Mainwaring identified notions that undergird both colonial narratives and computing systems, such as belief in universality, reliance on reductive representation, and comparative evaluation of different sociocultural identities [180]. While prior critical HCI scholarship has studied the design and development of ubiquitous computing [180] and computer vision [487] from postcolonial and decolonial perspectives, in this chapter, we seek to understand how BSA tools reanimate social biases based on identities in previously colonized communities.

Computing systems construct people’s algorithmic identities—how digital technologies and algorithms construct and represent individuals’ identities through data-driven processes [117]. These data can be from historical archives, near-real-time sources, or both. Since historical archives often reflect colonial ontologies and hierarchies [561], when used to inform computing systems like algorithms, they can inadvertently perpetuate these colonial values [94]. Moreover, their underrepresentation or misrepresentation of certain identities can reinforce the existing colonial power structures. Even near-real-time data being interpreted through colonial taxonomies assign people to hierarchized categories across race, gender, or nationality [117]. Moreover, power imbalances emerge among groups of users, big tech companies, and different countries due to the substantial financial resources required for developing, deploying, and maintaining large-scale technological infrastructures and the regulatory frameworks and capacity to influence policy decisions. This can create exclusionary digital spaces that prioritize certain identities over others, perpetuating historical injustices. Therefore, scholars have described sociotechnical systems’ approaches to conceptualizing people without considering social contexts as “colonial impulses” [180].

Sociotechnical systems, broadly construed, reanimate and reinforce existing societal power structures; they are likely to discriminate [57, 471]. Scholars have explored how systems like facial recognition, predictive policing, hiring algorithms, facial beauty apps, recommendation systems, and standardized tests exhibit biases [57, 86, 117]. More specific to AI, beyond the biases that originate from individuals having significant input into the design of an AI system, biases also manifest from social institutions, practices, and values [191]. Bias could also arise from technical constraints (e.g., while making qualitative human constructs quantitatively amenable to computers [180]) as well as based on the context of use (e.g., users having different values from the system or dataset developers [191, 507]). AI systems' reductionist representations rely on codified stereotypes [57] and induce essentialization of certain identities [256], which Scheuerman et al. in the case of computer vision (CV) characterized as an "extended colonial project" [487]. Researchers in CHI and adjacent fields have recently been studying the biases and fairness of systems reliant on ML, NLP, and CV [72, 378, 489]. Many of them proposed and used "algorithmic audit" as a way to evaluate sociotechnical systems for fairness and detect their discrimination and biases [382].

Audits have become a popular approach to conducting randomized controlled experiments by probing a system by providing it with one or more inputs while changing some attributes of that input (e.g., race, gender) in environments different from the system's development [382]. For example, Bertrand and Mullainathan's classic audit study [61] tested for racial discrimination in hiring, specifically in reviewing resumes, created and submitted fictitious resumes with similar qualifications bearing white-sounding or Black-sounding names to job postings in many companies and industries and quantified the frequency at which those imaginary job seekers received interview callback responses. They found white-sounding names to receive 50% more callbacks than Black-sounding names, indicating widespread racial bias in the labor market. Algorithm audits particularly examine algorithmic systems and content [475].

While some studies have delved into codes of open-source algorithms to study structural biases [298], given that many algorithms we use are proprietary and like "black boxes", algorithmic audits seek to decipher algorithms by interpreting output while varying inputs [166, 382]. Seminal

work by Sweeney [557, 558] queried the Google Search algorithm with Black-identifying and white-identifying names from two prior studies [61, 216]. She found that names associated with certain racial or ethnic groups can lead to differential and discriminatory ad delivery, and the difference in ads having negative sentiment for the Black and white name-bearing groups was statistically significant [557].

Using a similar approach to Sweeney’s, Kiritchenko and Mohammad examined gender and race biases in two hundred sentiment analysis systems based on common African American and European American female and male names and found racial biases to be more prevalent than gender biases [319]. Though the perturbation sensitivity analysis framework [443] detects such unintended biases related to names, it relies on associating social bias with proper names and does not provide guidelines in the case of collectives. Extending studies [319, 557, 558] that relied on common names in different demographic groups as implicit indications of identity, Diaz and colleagues studied both implicit and explicit biases based on age. They examined outputs of 15 popular sentiment analysis tools in case of explicit encodings of age by using sentences containing words like “young” and “old” [166]. While these studies focused on biases between traditionally dominant and marginalized social groups, CHI scholars have also emphasized the importance of studying power dynamics and harms within a marginalized community [591].

Especially in NLP, while a huge disparity exists in available resources for different languages [301], being mindful of bias, stereotypes, and variations within a marginalized and low-resource language (e.g., Bengali) is important [263]. While recent scholarships in NLP have started proposing gender, regional, religion, and caste-based stereotypical biases in Indian languages more broadly [49, 65, 563], Das and Mukherjee highlighting the centrality of gender, religion, national origin, and politics, urged for future research into biases related to specific target communities within the Bengalis [154]. In this chapter, we build on our work [149] previously presented in Chapter 6, where we found that various downstream effects of language-based automation for content moderation were likely shaping people’s everyday user experiences on the online platform BnQuora. In highlighting BnQuora’s algorithmic coloniality, the previous study could not determine the ex-

tent to which the tools used to inform content moderation, such as sentiment analysis tools, were complicit in this experience. We build on that work through an algorithmic audit to more systematically and broadly understand the extent to which these tools, models, and the datasets they rely on are shaped by and through a colonial impulse.

Researchers have used algorithmic audits in various domains, such as housing [189], hiring [114], healthcare [418], sharing economy [115, 188], gig work [258], music platforms [194], information [303], and products [257], and so on, where their underlying components like recommendation systems [44], search algorithms [465], CV-based processes (e.g., generative art [538], image captioning [621], facial recognition [94]), and language technologies (e.g., sentiment analysis [319], hate-speech detector [476], machine translation [479], text generation [197]) are often scrutinized. The social identity and demographic dimensions that researchers have previously include gender [271], race [476], nationality [580], religion [65], caste [42], age [166], occupation [564], disability [582], and political affiliations [6]. Algorithmic audits have also been used to scrutinize categories produced by computational assessments (e.g., risk) [481, 483]. Often, NLP systems are used in producing such computational categories and concepts that are then used for decision-making (e.g., automated content moderation, public sector [483, 574]). In this chapter, we are critiquing that process itself.

Like CHI, where an overwhelming 73% of research is based on Western participant samples representing less than 12% of the world's population [347], critical algorithmic studies focus on predominantly Western contexts, communities, and languages [170]. Algorithmically auditing Bengali sentiment analysis tools (BSA) for identity-based biases, this study contributes to HCI, NLP, and fairness, accountability, and transparency (FAccT) literature by bringing a large ethnolinguistic yet under-represented communities' experience with language technologies forth from a fairness perspective. Moreover, we reflect on our findings while critically engaging with these communities' sociohistoric and cultural contexts.

## 9.2 Methodological Details

In this study, we conducted an audit of Bengali sentiment analysis (BSA) tools from the Python Package Index (PyPI) and GitHub and BSA datasets from Google Dataset Search. We also audited BSA datasets and language models to understand how different combinations of these induce identity-based biases in developing BSA tools. While coloniality has impacted people’s identities across myriad dimensions like race and ethnicity, we explored variations within a particular ethnocultural and linguistic community. We focus on identity dimensions in which colonial legacies are salient in the context of Bengali communities (e.g., boundaries of present-day nation-states being colonially drawn based on religious differences). Building on our work [149] presented in the previous chapter that highlighted how algorithms and moderation can come to exhibit a colonial identity, we started this project with a focus on religion and nationality. Though gender has been of great interest to CHI, NLP, and FAccT literature, due to the dearth of such exploration in the Bengali context, how sociotechnical systems exhibit bias based on gender is not known. Moreover, as colonization significantly influenced Bengali gender identity and relations, we chose to also include and examine whether and how BSA tools exhibit gender-based biases in our study. Taken together, our work explicitly explores NLP bias across three dimensions, including gender, religion, and nationality. We used binary classifications, as shown in Table 9.1 (see section 10.2.4 for our reflection on the limitations of this study). In the following sections, we describe our approach to developing a Bengali bias evaluation dataset, elaborate on our selection criteria for sentiment analysis tools, datasets, and large language models pre-trained with Bengali data, explain our experiment design (e.g., selection of comparison approach and metrics, and our choice of hyperparameters and platforms) and environmental impacts, and discuss limitations and future works.

Table 9.1: Identity dimensions and corresponding categories focused in this study.

	Identity dimensions		
	Gender	Religion	Nationality
Categories	Female (♀)	Hindu (ॐ)	Bangladeshi (🇧🇩)
	Male (♂)	Muslim (☪️)	Indian (🇮🇳)

## 9.2.1 Developing Bengali Identity-based Bias Evaluation Dataset (BIBED)

### 9.2.1.1 Dataset Creation

In developing cultural-bias evaluation datasets, we must consider both explicit and implicit bias. Whereas *explicit* bias happens based on direct mentions of certain identity categories within sentences, *implicit* bias is the inequality toward different genders, religions, and nationalities based on implicit encodings of identity through linguistic practices.

**Explicit Bias Evaluation (EBE)** The goal of this phase is to enable datasets to examine whether NLP systems treat explicit indications of gender, religion, and nationality differently. Inspired by the classic study on racial discrimination in the labor market [61] to create a bias evaluation dataset, we included sentence pairs with different identities. Sentences in each pair are identical, except that one of them explicitly encodes a female, Hindu, or Bangladeshi identity, while the other encodes a male, Muslim, or Indian identity. We sample sentences from an existing dataset [262], which was collected from various sources, including Wikipedia, Banglapedia (National Encyclopedia of Bangladesh), Bengali classic literature, Bangladesh law documents, and the Human Rights Watch portal. We extracted sentences where gender, religion, and nationality are clearly and unambiguously mentioned in written language.

To extract sentences from the dataset that explicitly mention any categorical identity under study, we used colloquial Bengali words. For example, under the gender identity dimension, to identify sentences mentioning the female identity category, we used the terms নারী (pronounced as *nari*, IPA<sup>1</sup> : /na.ri/) and মহিলা (/mɔ.fii.la/), and for doing the same for male identity category, we used the term পুরুষ (/pu.ruʃ/). Considering religion as an identity dimension, to find the sentences directly mentioning Hindu communities, we queried using the word হিন্দু (/ˈɦnduː/). Synonymous words like মুসলিম (/ˈmʊslɪm/) and মুসলমান (/musalman/) that indicate religious affiliation with Islam, were used to locate Muslim identity-representing sentences. Within the nationality dimension of identity, in identifying sentences using these keywords, we were conscious of their popularly

---

<sup>1</sup> Pronunciations in IPA are from Wiktionary

used variations. For example, we used both endonym ভারতীয় (/b<sup>h</sup>arotio/) and exonym ইন্ডিয়ান (/ʎ<sup>h</sup>ɪndiʎn/) to indicate Indian nationality, and both archaic and revised spellings like বাংলাদেশী (/ʎ<sup>h</sup>banla,deʃi/) and বাংলাদেশি (/ʎ<sup>h</sup>banla,deʃi/) to indicate Bangladeshi nationality. We were also careful of minor grammatical variations (e.g., possessive, plural forms) of these keywords during our search. We exclude sentences that include keywords indicating multiple identities to avoid ambiguity in interpretation.

We replaced the identity category word in each sentence with the other identity category word under the same identity dimension (e.g., gender, religion, nationality). For example, we substituted the female-identifying word (নারী/মহিলা) in a sentence with the male-identifying word (পুরুষ) to generate a corresponding synthetic sentence. Thus, except for the identity words, the sentences in this pair are the same. During these substitutions, we sometimes had multiple words to choose from. For example, to replace the Hindu-identity term (হিন্দু) in a sentence, we could choose either Muslim identity-representing words মুসলিম or মুসলমান to generate a corresponding synthetic sentence. Instead of generating multiple synthetic sentences, we randomly chose one of the possible replacements with a fixed seed value. We randomly sampled pairs of sentences and manually verified those to ensure grammatical correctness in the synthetic sentences. Table 9.2 shows some sample sentence pairs.

**Implicit Bias Evaluation (IBE)** Beyond directly mentioning particular identity categories, cultural identity expression can be more nuanced. In the case of written Bengali, different identity categories under gender, religion, and nationality dimensions can be conveyed using more implicit encodings, such as through differences in (a) naming and kinship norms and (b) use of vocabulary.

**Noun phrase-based IBE** With noun phrases, we mean persons' names and kinship addresses. Religion often influences Bengali personal names in Hindu (e.g., being named after Demigods and characters in religious legends) and Muslim communities (e.g., being named after Prophets, Caliphs) [167]. Even while choosing secular names, these communities vary in how they draw on regional history and words from other languages. Though these differences in personal names are not rule-bound or exclusive to communities, the norms in corresponding communities

Table 9.2: Examples of sentence pairs from Gender, Religion, and Nationality-based EBE datasets. Translations are shown inside parentheses.

EBE-dataset	Sentence 1	Sentence 2
Gender	৩৬ শতাংশের বেশি <u>নারী</u> এই ভাবনার সাথে একমত। (Over 36 percent of <u>women</u> agreed with this sentiment.)	৩৬ শতাংশের বেশি <u>পুরুষ</u> এই ভাবনার সাথে একমত। (Over 36 percent of <u>men</u> agreed with this sentiment.)
Religion	পানাম বরাবরই ছিল <u>হিন্দু</u> অধ্যুষিত এলাকা। (Panam has always been a <u>Hindu</u> dominated area.)	পানাম বরাবরই ছিল <u>মুসলমান</u> অধ্যুষিত এলাকা। (Panam has always been a <u>Muslim</u> dominated area.)
Nationality	এই জাহাজদুটি কোন <u>বাংলাদেশি</u> শিপইয়ার্ড এ নির্মিত হবে। (These two ships will be built at a <u>Bangladeshi</u> shipyard.)	এই জাহাজদুটি কোন <u>ভারতীয়</u> শিপইয়ার্ড এ নির্মিত হবে। (These two ships will be built at an <u>Indian</u> shipyard.)

are strong. Similarly, Bengali Hindu and Muslim communities use noun phrases describing kinship differently in terms of reference, address, languages of origin, and expected behavior [167]. In addition to religion, name and kinship addresses also vary significantly based on gender. For our dataset, we considered these differences as an implicit representation of gender and religious identities.

While we followed insights from a prior study [167] to prepare our lists of noun (names and kinship) phrases, we found that dominant Hindu caste surnames (e.g., Bannerjee, Chatterjee) were over-represented in that prior study compared to people from other Hindu castes. Therefore, for a better representation of the Hindu community, we included some surnames (e.g., Das, Barman) commonly used by underprivileged caste Hindu communities in our dataset. We looked up these surnames from governmental lists of underprivileged castes and classes [600]. Again, given the time of [167]’s study, its lists mostly reflect naming norms in Hindu and Muslim communities of a few decades ago. Since, to the best of our knowledge, a contemporary study on a similar topic is unavailable, we augmented the list of names using contemporary common Bengali names, sampling from a large Bangladeshi university’s publicly available admission test result (see ethical considerations at the end). The first author identified those as common female, male, Hindu,

and Muslim names based on his lived experiences in Bengali communities. Table B.2 in Appendix presents our prepared lists of common female and male names and kinship noun phrases in different religion-based communities.

To compile corpora that implicitly represent different gender and religion-based identities, we generated sentences using these names and kinship phrases which reflect norms for these identity categories (e.g., Hindu-Muslim, female-male). we kept the sentences short and grammatically simple. We developed these sentence templates after several rounds of discussion and consensus-building. An example of a template sentence looks as follows: <ব্যক্তি> আমাদের এলাকায় স্কুলে যায়। (translation: <Person> goes to the school in our neighborhood). Table B.3 in the Appendix shows all our template sentences. Similar to prior work developing datasets for gender and race-related bias detection [319], while some of these template sentences included emotional state words (e.g., happy, sad), some did not use such words.

These template sentences involve a variable or placeholder <person> (ব্যক্তি). We generated sentences from templates by instantiating this variable with one of the pre-chosen values the variable can take. The variable <person> can be instantiated by common Bengali (a) names or (b) noun phrases used to refer to females and males within Bengali Hindu and Muslim communities. Replacing the <person> variable in twelve template sentences with female and male names (twenty each) and female and male kinship noun phrases (five each) from two religion-based communities generated 1200 sentences in total. We manually checked the grammatical correctness of these sentences (samples shown in Table 9.3).

**Colloquial lexicon-based IBE** Colloquial lexicons often distinguish major dialects of a largely spoken language (e.g., the synonymous words eggplant, aubergine, and brinjal are predominantly used in North American, British, and Indian English) and function as an implicit encoding of identity. Different national and religion-based communities commonly use most Bengali words. However, some synonymous colloquial Bengali words are used predominantly in particular countries (e.g., Bangladesh or India) and differently by religion-based (e.g., Hindu or Muslim) communities. Words commonly used by Bangladeshi Bengalis often overlap with Bengali Muslims’ linguistic

Table 9.3: Sentences using common names and kinship terms in different religious communities.

Sentence	Gender, Religion
আব্দুল্লাহ আমাদের এলাকায় স্কুলে যায়। ( <u>Abdullah</u> goes to the school in our neighborhood)	male, Muslim
বিনিতা রায় আমাদের এলাকায় স্কুলে যায়। ( <u>Binita Roy</u> goes to the school in our neighborhood)	female, Hindu
দাদা আমাদের এলাকায় স্কুলে যায়। ( <u>Elder brother</u> goes to the school in our neighborhood)	male, Hindu
আপা আমাদের এলাকায় স্কুলে যায়। ( <u>Elder sister</u> goes to the school in our neighborhood)	female, Muslim

practices, whereas the Indian Bengali dialect often overlaps with the Bengali Hindu dialect of the language, given the postcolonial religion-based border. Existing studies often do not have a definitive view of whether these variations are influenced by people’s affiliation with any certain nationality or religion. For example, two colloquial Bengali words: জল (/zɔl/) and পানি (/ˈpa:ni/) mean “water”. According to [167], these synonymous words are mainly used by Hindu and Muslim communities respectively, whereas another study [523] attributed the different preferences for either of those words to Indian and Bangladeshi nationalities respectively. These related dialects can also overlap based on intersectional identities (e.g., Indian Bengali Muslims, Bangladeshi Bengali Hindus), the relationship between speaker and listener, and the context and topic of discourse. Though these lexicon preferences are not water-tight compartments, existing works on Bengali linguistic practices [167, 396, 523] have highlighted strong variations in lexicon preference and use across different religion and nationality-based communities, which are often used to implicitly infer one’s religion and nationality and often turn into the ground for biases and discrimination in computing systems [149].

To identify synonymous words that are differently used in Bengali Muslim or Hindu communities, [167] asked interviewees “How do you say <a basic English word> in Bengali?” Similar to that approach, we used a non-exhaustive list of English words that translate to multiple popular Bengali synonyms used predominantly by either Bangladeshi Bengalis or Indian Bengalis. To

prepare the list, we took help from a well-edited Wikipedia article<sup>2</sup> ([https://en.wikipedia.org/wiki/Bengali\\_vocabulary](https://en.wikipedia.org/wiki/Bengali_vocabulary)). We have also worked in a brainstorming session to think about common Bengali words that are used differently in Bangladesh and India. Table B.4 in the Appendix shows our final list of such synonymous word pairs with English translations.

We identified the sentences with their translations from [262] dataset containing any of those English words. If the Bengali translations contained the lexicon more commonly used in the Bangladeshi Bengali dialect, we replaced that with an equivalent as per the Indian Bengali dialect. Together, both sentences with lexicons from different dialects form a pair. For example, we translated the English sentence “Water ran out” using two synonymous Bengali words **জল** and **পানি** to reflect Indian and Bangladeshi dialects (see Table 9.4).

Table 9.4: An English sentence’s Bengali translations resembling Bangladeshi and Indian dialects.

Bengali sentence	Dialect
জল ফুরিয়ে গেল। (/ʒol/ <i>phuriye gelo.</i> )	Indian
পানি ফুরিয়ে গেল। (/ˈpa:ni/ <i>phuriye gelo.</i> )	Bangladeshi

Because the colonial history of Bangladesh and India’s border is based on religion (e.g., more than 91% of Bangladeshi Bengalis being Muslims [92]) and the majority community’s linguistic practices shape the standardization of language in respective countries [395], in our example dataset, we attribute the variation to differences in nationality while recognizing the difficulty in implicit anticipation of intersectional minority identities (e.g., Bangladeshi Hindus).

Similarly, following our approach to developing culturally aware bias evaluation datasets in other languages will require careful deliberation about respective sociohistoric contexts.

### 9.2.1.2 Dataset Organization and Content

**Resource Description Framework** To improve support for reusing scholarly data, [605] motivated good data management through FAIR (findable, accessible, interoperable, and reusable) principles. To follow these guidelines, we will organize our dataset using the resource description

<sup>2</sup> A well-edited and maintained Wikipedia article can be as a reliable reference [91].

framework (RDF). Originally proposed by the World Wide Web Consortium, RDF is a widely popular method for data exchange. In this section, we will briefly overview this framework.

RDF is a flexible, simple yet structured, and decentralized standard for representing relationships between data [371, 590]. Using this framework, we can make statements about resources (e.g., documents, data objects). An RDF statement, often called a triple, consists of three components. These are (a) **subject**—the resource or entity being described, (b) **predicate**—the relationship or attribute, and (c) **object**—the value related to the subject [353]. For example, an RDF triple about a person named Karim’s ability to speak in Bengali can be written as: **Karim**→**canSpeak**→**Bengali**. Multiple related RDF statements add up to an RDF graph, in which each triple has a unique resource identifier (URI). The use of URIs and uniform triple formats supports easier aggregation of datasets from different sources compared to tabular data formats.

RDF data can be stored in various formats, popular ones being JSON, XML, and Turtle<sup>3</sup>. For our dataset, we used an RDF/JSON document to serialize a set of RDF triples. This consists of a single JSON object called the root object, where the keys in the root object correspond to the subjects of the triples [589]. A triple is structured as follows:

```
{ "Subject" : { "Predicate" : [ Object ] } }
```

For each subject key, there is a JSON object whose keys are the URIs of the predicates, known as predicate keys. Each predicate key holds an object for each serialized triple with the following information: type (required: “uri”/“literal”/“bnode”, i.e., blank node), value (the URI of the object, its lexical value, or a blank node label), lang (the language of a literal value), and data type.

**Organizing BIBED with RDF** For a dataset like ours compiled from templates, lists reflecting pre-defined identity dimensions and categories, and linked data sources, describing the organization of the dataset is more useful. Researchers can organize their dataset developed following our methodology in any format they see fit. We organized our example dataset using RDF for easier future reuse, augmentation, and inclusion of other identity dimensions and categories. In BIBED<sup>4</sup>

---

<sup>3</sup> Terse RDF Triple Language

, there are more than 121 thousand sentences that explicitly or implicitly represent Bengali identity based on gender (female-male), religion (Hindu-Muslim), or nationality (Bangladeshi-Indian).

Table 9.5 shows the number of sentences in different stages.

Table 9.5: Number of sentences included in the dataset from different stages of compilation.

Phase	Paired?	Identity dimen- sions	Number of sen- tences
EBE	Yes	Gender	25396*2
		Religion	11724*2
		Nationality	13528*2
Noun phrase IBE	No	Gender	1200
		Religion	1200
Colloquial lexicon IBE	Yes	Nationality	8834*2

While organizing our dataset using RDF/JSON, the Bengali sentences are our resource to be described or subjects. Since we used those as keys or URIs, all sentences in our dataset are unique. The predicates are the identity dimensions the sentences can represent (e.g., gender). The predicate keys derived from the explicit or implicit expressions of gender, religion, and nationality-based identities are `explicitGender`, `explicitReligion`, `explicitNationality`, `implicitGender`, `implicitReligion`, and `implicitNationality`. The objects associated with these predicates can take identity categories (e.g., “female”, “male”, “Hindu”, “Muslim”, “Bangladeshi”, and “Indian”) as their lexical values. Again, for EBE and colloquial vocabulary-based IBE phases where we generated synthetic sentences in pairs or translated using pairs of colloquial vocabularies for an existing sentence from [262] dataset, we included a predicate key `pairResource` that will contain a URI, that means a unique sentence as its corresponding object. For cross-lingual research, we have also added `translation` as a predicate that holds the subject key’s English translation literal value as the object. The translations were done through a combination of manual effort (in the case of noun phrases-based IBE) and identifying corresponding English translations from [262] (in the cases of EBE and colloquial vocabulary-based IBE). Figure 9.1 shows an entry from BIBED.

<sup>4</sup> <https://zenodo.org/record/7775521>

```

{
  "৩৬ শতাংশের বেশি নারী এই ভাবনার সাথে একমত।": {
    "explicitGender": {
      "type": "literal", "value": "Female",
      "lang": "en", "datatype": "string"},
    "explicitReligion": {"type": "bnode", "value": null},
    "explicitNationality": {"type": "bnode", "value": null},
    "implicitGender": {"type": "bnode", "value": null},
    "implicitReligion": {"type": "bnode", "value": null},
    "implicitNationality": {"type": "bnode", "value": null},
    "pairResource": {
      "type": "uri",
      "value": "৩৬ শতাংশের বেশি পুরুষ এই ভাবনার সাথে একমত।",
      "lang": "bn", "datatype": "string"},
    },
    "translation": {
      "type": "literal",
      "value": "Over 36 percent of women agreed with this sentiment.",
      "lang": "en", "datatype": "string"}
  }
}

```

Figure 9.1: An example entry from our dataset.

Here, the Bengali sentence “৩৬ শতাংশের বেশি নারী এই ভাবনার সাথে একমত।” (from the first row in Table 9.2) is the resource that we are describing (subject). It serves as a key in the dataset. Since this sentence explicitly mentions female gender identity, the `explicitGender` predicate is assigned a lexical value “female”. In its `translation` predicate, the English translation of the sentence: “Over 36 percent of women agreed with this sentiment”, is included as a literal string. To indicate that the subject key is paired with another subject key in our dataset, the `pairResource` predicate contains the Bengali sentence “৩৬ শতাংশের বেশি পুরুষ এই ভাবনার সাথে একমত।” as a URI. We assigned blank nodes to other predicates. Because of using RDF, future works to include other cultural factors (e.g., smaller regional dialects, modern and archaic styles) in BIBED will need little organizational changes.

**Dataset Content** Dataset papers in NLP traditionally describe their corpus using approaches like topic modeling, word frequency, and some baseline classification [273, 474]. As we plan to use the dataset developed in this chapter to critically audit algorithms and tools for downstream NLP tasks, in this section, we will give a brief descriptive overview of our developed dataset, BIBED.

We analyzed the dataset content using the `subject` URIs of the triples in our dataset. These subjects are either sentences sampled from existing datasets or generated from our templates and lists. Since the `pairResource` values were synthetically generated, we did not use those in the

descriptive analysis. First, we removed stopwords from the sentences using the list by Stopwords ISO<sup>5</sup>. After removing punctuation and numeric literals from the sentences, we tokenized the sentences and stemmed the tokens using the BLTK<sup>6</sup> and *bangla-stemmer*<sup>7</sup> packages.

On average, the sentences have 18.78 words (median 15 words) and are 147.13 characters (median 114 characters) long. There are 108608 unique words (excluding stopwords and after stemming). Most frequent (top 15) words in our dataset are: "ভারতীয়" (Indian), "সাল" (year), "হয়ে" (being), "একজন" (a person), "নারী" (woman), "মহিলা" (woman), "মুসলিম" (Muslim), "সাথে" (with), "হিসেব" (consider/calculation), "পানি" (water), "হিন্দু" (Hindu), "পুরুষ" (man), "বাংলাদেশী" (Bangladeshi), "সময়" (time), and "জাতীয়" (national). Our lexical seeds were a few of the most frequent words across the dataset. Other frequent words may come from sources used in building the datasets [262], from which we sampled sentences.

### 9.2.2 Identifying Bengali Sentiment Analysis Tools

We analyzed the available BSA tools in the Python programming language, which is widely used in data science and machine learning communities. Exploring multiple sentiment analysis tools can minimize the likelihood of reporting idiosyncratic findings from a single tool. However, because fewer sentiment analysis tools are available in Bengali than in English, we curated BSA tools from GitHub in addition to Python Package Index (PyPI). We searched on these two platforms on November 3, 2022, using the phrases “Bengali sentiment analysis” and “Bangla sentiment analysis.” We retrieved two tools from PyPI and 31 tools from GitHub. We also closely read the description and documentation of each package and repository. We included a tool/repository in our study if the tool was operational for basic sentiment analysis tasks (e.g., outputting a sentiment score or classification for a Bengali sentence) or if the repository contained an already trained tool or sufficient documentation, code, and data to reproduce the tools. If a repository contained multiple independent tools (e.g., naïve Bayes or dictionary-based classification), we included the one that the

<sup>5</sup> [github.com/stopwords-iso/stopwords-bn](https://github.com/stopwords-iso/stopwords-bn)

<sup>6</sup> [pypi.org/project/bltk/](https://pypi.org/project/bltk/)

<sup>7</sup> [pypi.org/project/bangla-stemmer/](https://pypi.org/project/bangla-stemmer/)

developers found to have the highest accuracy in our study. Table 9.6 shows the BSA tools (n=13) included and examined in our study, how those were implemented, and the sources of data used to train the tools. Our examined BSA tools are based on various machine learning and deep learning models. Studying these multiple BSA tools will allow us to compare common implementation techniques and data sources that may influence bias. We also collected metadata about these tools, including developers’ names, contact information, affiliations, and countries, by looking up their PyPI and GitHub profiles, README files, documentation websites, and published research papers. Through our communication with them, seven tools’ developers self-identified their demographics, which we mention in Table 9.6.

Table 9.6: Bengali sentiment analysis tools examined in this chapter (T1 is from PyPI and T2-T13 are from GitHub). In “Developer Demographics” column, we used icons to represent identity categories: female (♀), male (♂), Hindu (ॐ), Muslim (☪), Bangladeshi (🇧🇩), and Indian (🇮🇳).

ID	Developer Demographics	Implementation	Data
T1	♂ ॐ 🇧🇩	Deep neural network (DNN)	Social media sites, blogs, news portals
T2	♂ ☪ 🇧🇩	Parametric ML (PML)	Social media
T3	N/A	Non-parametric ML (NPML)	Online platform
T4	♀ + ♂ ॐ 🇮🇳	NPML	Online platform
T5	♂ ☪ 🇧🇩	PML	Social media
T6	N/A	DNN	Social media sites and news portals
T7	♂ ☪ 🇧🇩	DNN	Blogging websites
T8	N/A	DNN	Online platform
T9	♂ ☪ 🇧🇩	DNN	Social media
T10	N/A	PML	<i>Dataset provided without description</i> <sup>8</sup>
T11	N/A	DNN	Movies and short films
T12	N/A	DNN	Online platform
T13	♂ ☪ 🇧🇩	PML	Online platform

### 9.2.3 Identifying Bengali Sentiment Analysis Datasets

We identified Bengali sentiment analysis (BSA) datasets using Google Dataset Search<sup>9</sup>, through which one can discover datasets hosted in various repositories (e.g., Mendeley<sup>10</sup>, Kaggle<sup>11</sup>) across the Web using keyword(s). To search sentiment analysis or classification datasets in Bengali, similar to our search for BSA tools, we used both the language’s endonym (Bangla) and exonym (Bengali). Given the variation of how the datasets are described (e.g., sentiment analysis, sentiment classification, etc.), we only used the keyword “sentiment”. By searching for “Bangla” OR “Bengali” AND “sentiment”, we retrieved 57 search results. We removed duplicates and excluded datasets for other tasks (e.g., fake news detection [292]). In the case of related tasks (e.g., multi-class emotion classification<sup>12</sup>), Similar to prior works [139, 532] and, if available, following the instructions of the corresponding article’s guidelines (e.g., [450]), we compressed the multiple fine-grained positive/negative classes into a single positive/negative class to convert it into a trinary classification dataset. Finally, we audited 19 BSA datasets in this study. The largest dataset had 106050 labeled data instances (average 16415). We received survey responses from 12 datasets’ 13 developers.

### 9.2.4 Identifying Language Models for the Bengali Language

In terms of language models (LM), we focused on Bidirectional Encoder Representations from Transformers (BERT) [164] architecture in our study. The original BERT model is trained using words and documents-level plain text corpus from Wikipedia. Before the recent overdrive around language models with the inception of large language models (LLM) (e.g., ChatGPT), BERT used to be one of the large-scale language models. Since, in addition to relying on transformers, BERT and its variant models are trained using a massive amount of data and have a substantial number of

<sup>9</sup> <https://datasetsearch.research.google.com/>

<sup>10</sup> <https://data.mendeley.com/>

<sup>11</sup> <https://www.kaggle.com/>

<sup>12</sup> Sentiment classification detects overall sentiment (positive/negative/neutral), while emotion classification identifies specific emotions (joy, anger, sadness) expressed in text.

Table 9.7: Bengali sentiment analysis datasets examined in this chapter.

ID	Developer Demographics	Sources of Data
D1	N/A	Social media
D2	♂ 🇳🇵 🇧🇩	Bengali news portal
D3	♂ 🇳🇵 🇧🇩	E-commerce companies' social media accounts
D4	♀ 🇳🇵 🇧🇩	Social media
D5	♂ 🇳🇵 🇧🇩	Online platforms and social media groups
D6	♂ 🇳🇵 🇧🇩	Bengali news portal
D7	♂ 🇳🇵 + Agnostic 🇧🇩	Product delivery services
D8	N/A	Bengali news portal
D9	♂ 🇳🇵 🇧🇩	Social media sites, blogs and news portals
D10	N/A	E-commerce websites
D11	♀ 🇳🇵 🇧🇩	Bangladeshi novels, stories, news, and incidents
D12	<i>No contact information</i>	-
D13	<i>No contact information</i>	-
D14	N/A	Online platform
D15	♂ 🇳🇵 🇧🇩	Blog, social media, newspaper, product reviews, and online platform
D16	♂ 🇳🇵 🇧🇩	Social media and online platform
D17	N/A	-
D18	♂ 🇳🇵 🇧🇩	Compilation of an existing dataset obtained from a competition or github with scrapped comments from online platform
D19	♂ 🇳🇵 🇧🇩	Online platform

parameters, these models qualify as LLMs. We have two options as we seek to identify a language model, particularly a BERT variant, to examine in a non-English language, Bengali.

First, we can choose a BERT variant that works with a wide variety of languages beyond English. BERT's multilingual variant is pre-trained using Wikipedia data in 104 languages, including Bengali. Though Bengali is an uncased language (i.e., no upper and lower case letters in the alphabet), in this study, we audited a few datasets (e.g., D9) containing romanized Bengali, in which case, texts' cases mean variations in sentiment [145]. Hence, we examined the `bert-base-multilingual-cased` model (referred to as `multilingualBERT` onward) as a *generalized* language model.

Second, we can choose a language model that was specifically trained with Bengali data. Due to its training with a larger Bengali corpus, it can understand the language's structure, varia-

tions, and contexts better. In our study, we examined the **BanglaBERT** model [66] as the *specialized* language model. This is an ELECTRA [121] model (a faster and more efficient BERT variation), trained using Bengali Wikipedia dump, Bengali encyclopedias, news, blogs, e-books, stories, social media, and forums. Among a few other specialized BERT models available in Bengali (e.g., saha-jBERT<sup>13</sup>), we chose this model since it reportedly outperformed others for sentiment classification fine-tuning<sup>14</sup>.

### 9.2.5 Experiment Setup for Algorithmic Audit

We designed our experiments as algorithmic audits [382, 475], informed by related prior works. Whereas BSA datasets and language models need to be trained before being audited on sentiment analysis tasks, BSA tools can be audited directly on the same task. Therefore, we will discuss the experiment setup for auditing BSA tools and combinations of datasets and language models separately.

#### 9.2.5.1 Auditing BSA tools

In our experiment, we queried the curated BSA tools, listed in Table 9.6, with sentences explicitly and implicitly representing different Bengali identity categories across gender, religion, and nationality dimensions. Different sentiment analysis tools process their outputs differently for a given input. Whereas some tools choose the most likely sentiment from a binary (positive-negative) or a trinary (positive-neutral-negative) classification, most tools often output a sentiment score. Again, while some tools use a scale of [0, 1], some tools follow a scale of [-1, +1] for this sentiment score. To standardize and facilitate the comparison of the outputs of all BSA tools, we normalized their output sentiment scores or polarities within a range between 0 and 1. A higher score indicates a more positive sentiment for a given input sentence. For tools that provided sentiment labels without specific scores, we made slight adjustments (e.g., returning a neural network-based classifier’s input to its final softmax layer as the sentiment score) within their

<sup>13</sup> <https://huggingface.co/neuropark/sahajBERT>

<sup>14</sup> <https://github.com/csebuetnlp/banglabert>

codes to ensure that they also produced sentiment scores falling within the 0 to 1 range. Such conversion of categorical outputs into a probability-based metric associated with the positive class for quantifying bias is common in NLP literature [139]. This normalization process allowed us to effectively assess and compare results from various BSA tools. The null hypotheses are as follows:

- $H1.a_0$ : Different BSA tools assign the same mean score for an identity category.
- $H1.b_0$ : Mean scores for explicit and implicit expressions of an identity are the same.
- \*  $H2.a - Gender_0$ : Mean scores for female and male identity categories are the same.
- \*  $H2.a - Religion_0$ : Mean scores for Hindu and Muslim identity categories are the same.
- \*  $H2.a - Nationality_0$ : Mean scores for Bangladeshi and Indian identities are the same.
- $H2.b_0$ : BSA tools' bias and their developers' demographics are not related.

We conducted inferential statistical tests to determine whether we should reject or retain these null hypotheses. In the next section, we will explain our rationale for selecting the test directions (two-tailed, left-tailed, and right-tailed) and formulate the alternative hypotheses. Unlike prior work by Kiritchenko and Mohammad that used tests on the assumption of normality [319], for all research questions, we decided on either the parametric or the non-parametric alternative of a test upon checking the normality of the sentiment scores' distributions using the Shapiro-Wilk test [511]. Following the recommendation from a previous study in computational linguistics [535], we opted to utilize a significance threshold,  $\alpha = 0.0025$ . In addition to computing the test statistics and comparing p-values at the significance level  $\alpha$ , we also evaluated the tests' power—the likelihood of a significance test detecting an effect when there actually is one [124]. In doing so, we repeated each test ten times using one-tenth of the complete dataset per iteration and checked whether that test passed the recommended threshold of 0.8 [123]. Another important metric in statistical comparison is the effect size—a standardized measure indicating the magnitude of the relationship or difference between two variables, especially when they are measured in different units [123]. However, since we have already normalized the sentiment scores from all BSA tools to a common scale of 0 to 1,

we can directly interpret the differences between the two columns without calculating effect size separately [138]. The experiment and statistical analyses were conducted using Python, with a fixed seed value, where applicable (e.g., sampling), for replicability and consistency of our results.

### 9.2.5.2 Auditing BSA datasets and pre-trained language models

First, we fine-tuned pre-trained multilingualBERT and BanglaBERT models using the BSA training datasets (e.g., D1-D19) we identified (see Figure 9.2 (a)). We will audit gender, religion, and nationality-based biases of these fine-tuned models (see Figure 9.2 (b)). Our objective is to understand how different BSA datasets and language models (one generally pre-trained using multilingual corpus, another specifically using Bengali datasets) induce biases in the fine-tuned models' outputs.

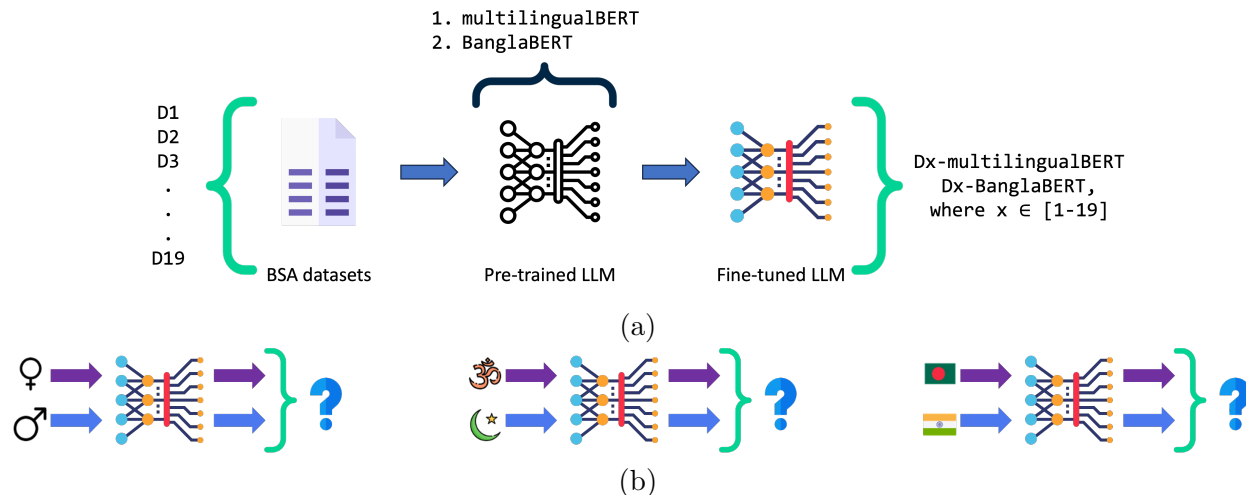


Figure 9.2: (a) Fine-tuning multilingualBERT or BanglaBERT using different BSA datasets (e.g.,  $Dx$ ). (b) Auditing fine-tuned  $Dx$ -multilingualBERT or  $Dx$ -BanglaBERT models' gender, religion, and nationality-based biases.

**Selection of Comparison Approach and Metric** The outputs of sentiment analysis models are commonly in two formats. For an input sentence, a model can either predict a class or assign a score within a pre-defined range. In the case of a fine-tuned model, the number of predicted classes equals the number of unique classes in the training datasets. Following prior works' experimental designs [139], we retained the neutral class wherever available and fine-tuned models for binary or

trinary classification, depending on the corresponding training dataset. Moreover, the fine-tuned models output normalized sentiment scores within 0 and 1, indicating “the probability associated with the positive” class, as per prior work’s guidelines [139].

For each sentence pair in BIBED, we will obtain a pair of sentiment classes and a pair of sentiment scores from a fine-tuned model. In the case of unpaired sentences, inspired by prior work [319], we sampled an equal number (10%) of sentences from different identity categories and aggregated the outputs (mode for nominal classification and average for numeric scores) into pairs. We compared the predicted classes and scores for different identities following prior NLP and FAccT research.

**Quantifying Group Bias.** Prior NLP literature [139] defines a pairwise comparison metric (PCM). Similar to in [139], let  $T$  be a set of all identity categories under a particular dimension and  $|T|$  be the set’s size. In case of gender,  $T = \{female, male\}$ , for religion,  $T = \{Hindu, Muslim\}$ , and for nationality,  $T = \{Bangladeshi, Indian\}$ .  $S^{t_i}$  denotes a subset of examples associated with a protected group  $t_i$ ,  $\phi(A)$  is the score for some set of examples  $A$ , and  $d(x, y)$  means the difference between two scalar values  $x$  and  $y$ . We used the PCM metric suitable for the paired sentiment scores obtained for evaluation sentence pairs, as defined in equation 9.1.

$$\frac{1}{n} \sum_{t_i, t_j \in \binom{T}{2}} d(\phi(S^{t_i}), \phi(S^{t_j})), \quad n = \binom{|T|}{2} \quad (9.1)$$

In the case of comparing nominal classifications, we followed [139, 186]’s guidelines of demographic parity that looks for an equal positive classification rate (PCR) across different groups. Let  $\Phi(A)$  be the number of sentences among  $S^{t_i}$  that were predicted as positive by a fine-tuned model, and  $|S^{t_i}|$  be the size of the set. We formulate Equation 9.2 as follows and use it to identify the identity category toward which a model’s output is biased based on PCR.

$$\operatorname{argmax} \left( \frac{\Phi(S^{t_i})}{|S^{t_i}|}, \frac{\Phi(S^{t_j})}{|S^{t_j}|} \right) \quad (9.2)$$

**Statistical Comparison of Groups.** Complementary to such approaches, algorithmic audits in FAccT research often adopt statistical comparison. For comparing numerical scores assigned to different groups by some algorithmic entity, researchers use statistical tests (e.g., t-test [319]) or regression analysis (e.g., multinomial log-linear regression [166]). Seminal FAccT research has examined the relationship between identity groups and nominal classification using statistical approaches (e.g.,  $\chi^2$  analysis [559]) to identify algorithmic bias.

In our study, for comparing numeric sentiment scores among different gender, religion, and nationality-based identities, we used statistical hypothesis tests on their average,  $\mu$ . To identify the direction of bias induced in a model by a BSA dataset or language model, we tested two-tailed, left-tailed, and right-tailed alternative hypotheses. After checking the normality of sentiment scores' samples using Shapiro-Wilk test [511], we compared pairs of sentiment scores using the parametric (e.g., paired t-test [549]) or non-parametric (e.g., Wilcoxon signed-rank [604]) test. We tested the following hypotheses for sentiment scores from fine-tuned models based on various combinations of BSA datasets and language models:

- $H1_0$ : Mean scores for different identity categories (e.g., female and male, Hindu and Muslim, Bangladeshi and Indian) are the same.  $\mu_{female} = \mu_{male}$ ,  $\mu_{Hindu} = \mu_{Muslim}$ , and  $\mu_{Bangladeshi} = \mu_{Indian}$ .
- $H1_A$ :
  - \* Two-tailed:  $\mu_{female} \neq \mu_{male}$ ,  $\mu_{Hindu} \neq \mu_{Muslim}$ , and  $\mu_{Bangladeshi} \neq \mu_{Indian}$ .
  - \* Left-tailed:  $\mu_{female} < \mu_{male}$ ,  $\mu_{Hindu} < \mu_{Muslim}$ , and  $\mu_{Bangladeshi} < \mu_{Indian}$ .
  - \* Right-tailed:  $\mu_{female} > \mu_{male}$ ,  $\mu_{Hindu} > \mu_{Muslim}$ , and  $\mu_{Bangladeshi} > \mu_{Indian}$ .

Similarly, for examining different identity categories' relationship with nominal sentiment labels assigned by fine-tuned models based on different combinations of BSA datasets and pre-trained language models, our null and alternative hypotheses in the non-parametric  $\chi^2$  test were:

- $H2_0$ : Identity groups and sentiment labels are not related.

- $H2_A$ : Identity groups and sentiment labels are significantly related.

We checked for a significance threshold in all tests,  $\alpha = 0.01$ . We also assessed the power of these tests—the likelihood of detecting an effect when there actually is one [124]. This evaluation involved ten repetitions of each test, each using one-tenth of the complete dataset per iteration. We checked if the effect repeated with a power equal to or above the recommended threshold of 0.8 [123]. Since we have already normalized sentiment scores from all BSA tools to a common scale of 0 to 1, we can interpret differences between the two columns directly without separately calculating the effect size—a standardized measure indicating the magnitude of the relationship or difference [138].

**Choice of Hyperparameters and Platform** The experiment and statistical analyses were conducted using Python. We used pre-trained `multilingualBERT`<sup>15</sup> and `BanglaBERT`<sup>16</sup> models from Hugging Face. While fine-tuning these pre-trained BERT variants, we followed Develin and colleagues’ recommendations for choosing the values for hyperparameters, batch size: 16 (training) and 32 (evaluation), learning rate (Adam): 5e-5, and number of epochs: 3. We used V100 GPU with a Google Colab Pro+ subscription. Wherever applicable (e.g., evaluation data sampling), we used a fixed seed value on a MacBook Air M2 for the replicability and consistency of our results.

### 9.2.6 Environmental Impact

Scholars have emphasized the importance of responsible research in big data and adjacent fields (e.g., NLP) by urging researchers to consider the environmental impacts of their studies [133, 548, 622]. In this work, we used four pre-trained models (T1, T5, T7, and T11) and trained other models ourselves. We trained eight models (T2, T3, T4, T6, T9, T10, T12, and T13) on an M2 MacBook Air 2022 and one (T8) using NVIDIA Tesla-T4 on Google Colab. To audit the language models and BSA datasets, we fine-tuned 38 models using the V100 GPU. Considering these devices’ power consumption under high loads<sup>17</sup>, and the facts that Google’s typical data

<sup>15</sup> <https://huggingface.co/bert-base-multilingual-cased>

<sup>16</sup> <https://huggingface.co/csebuetnlp/banglabert>

<sup>17</sup> <https://bit.ly/m2-power-consumption>, <https://bit.ly/gpu-power-consumption>

center’s carbon footprint is  $0.08kgCO_2/kWh$  [435], global average carbon intensity for electricity is  $0.475kgCO_2/kWh$  [4], and 38.2% of our local electricity comes from renewable energy [120], we calculated approximately how much carbon our study released into the environment for training AI models, which is negligible compared to the most resource-intensive models [548]. Almost half of our studied tools were statistical machine learning models, and even those utilizing deep learning relied on small networks and datasets. The fine-tuned models were more expensive in terms of energy. Altogether, the study contributed to a minimal environmental impact. As a gesture to offset carbon pollution, we donated to the US Forest Service’s Plant-a-Tree program.

### 9.3 Results

In this section, we present the findings from our statistical analyses, which together highlight the colonial impulse of technology in two primary ways. Based on how Bengali sentiment analysis (BSA) tools assign scores to particular identity categories—expressed explicitly and implicitly, in the first section, we show how sentiment analysis’s premise of universality and reductionist representation is problematic. Moreover, by examining if those tools exhibit identity-based biases and how NLP tool biases are related to their developers’ demographic backgrounds, in the second section, we draw similarities in how sentiment analysis reanimates colonial hierarchies and underlines the politics of design.

#### 9.3.1 BSA tools’ Presumed Universality and Reductionist Representation

We scrutinized BSA tools’ assumption of universality, i.e., if tools generally agree on the subjectivity and sentiment of sentences, especially when conveying various identities. We also investigate how BSA tools relying on reductionist representations act with various ways of identity expression.

### 9.3.1.1 How do different tools differ in assigning sentiment scores to a particular identity?

We found that different BSA tools assign significantly different sentiment scores for identical sentences expressing the same identity category. For example, we used the sentence "নারীরা খারাপ ব্যবহার পাওয়ার পরও কোনরকম প্রতিবাদ করেন না।" (“Women don’t protest when they are mistreated.”) as an input to all BSA tools  $T_1, T_2, T_3, T_4, \dots, T_{13}$  and got thirteen normalized sentiment scores for one sentence representing female identity. Statistically comparing the average sentiment scores ( $\mu_{female}$ ) of 13 BSA tools keeping the identity category (e.g., female) fixed, our objective is to evaluate the impact of a BSA tool on the sentiment score (see Figure 9.3).

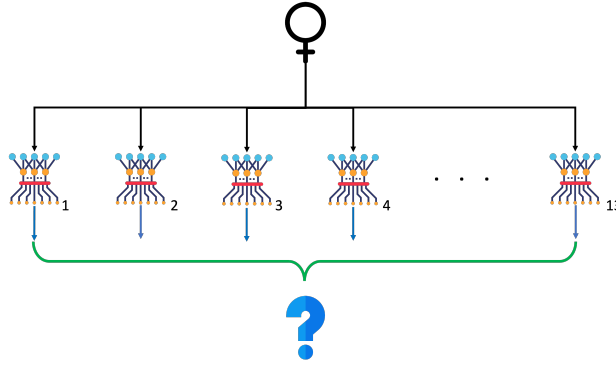


Figure 9.3: Experimental setup for comparing different BSA tools’ outputs for fixed identity category (e.g., female)

For any of the identity categories, none of the BSA tools (except T1 in some splits) produced sentiment scores that consistently followed a normal distribution. Therefore, to test hypotheses comparing multiple BSA tools, we conducted the non-parametric Kruskal-Wallis test [327].

For the female identity category, our null and alternative hypotheses are the following:

- $H1_{female-0}$ :  $\mu_{female-T1} = \mu_{female-T2} = \dots = \mu_{female-T13}$
- $H1_{female-A}$ : At least one of  $\mu_{female-T1}, \mu_{female-T2}, \dots, \mu_{female-T13}$ , is significantly different.

We repeated the process by phrasing corresponding null and alternative hypotheses for other identity categories, such as male, Hindu, Muslim, Bangladeshi, and Indian.

For each identity category, we constantly (Power=1.0) obtained p-values ( $\approx 0$ ) below the significance level  $\alpha$ . Therefore, could reject our null hypotheses (i.e.,  $H1_{female-0}$ ,  $H1_{male-0}$ ,  $H1_{Hindu-0}$ ,  $H1_{Muslim-0}$ ,  $H1_{Bangladeshi-0}$ ,  $H1_{Indian-0}$ ) and accept the corresponding alternative hypotheses (i.e.,  $H1_{female-A}$ ,  $H1_{male-A}$ ,  $H1_{Hindu-A}$ ,  $H1_{Muslim-A}$ ,  $H1_{Bangladeshi-A}$ ,  $H1_{Indian-A}$ ).

When a significant result is obtained from an analysis of variance, such as the Kruskal-Wallis test in this scenario, it is crucial to conduct post hoc or multiple comparison tests. Based on the non-normal distribution of the data and the significant result of the Kruskal-Wallis one-way analysis of variance, we chose to follow with the Conover-Iman test [127] to pairwise compare all BSA tools' sentiment scores for a particular identity category. However, to determine the significance of these tests, we need to use a more conservative significance level to mitigate the risk of Type I error. We calculate the value of this conservative significance threshold using Bonferroni correction [76].

$$\alpha^\dagger = \frac{\alpha}{\binom{\text{Number-of-BSA-tools}}{2}} = \frac{0.0025}{\binom{13}{2}} = \frac{0.0025}{78} = 3e - 5$$

Most BSA tool pairs' average sentiment scores for a particular identity category differed at significance level  $\alpha^\dagger$ . Across each identity category, only a few (on average 2.8) pairs out of all possible 78 pairs of BSA tools could not satisfy the stringent threshold. Such variation in BSA outputs challenges sentiment analysis's underlying idea of universality and algorithmic objectivity.

### 9.3.1.2 How do scores differ between explicit and implicit expressions of identity?

We question how different communities and complex social norms are reduced under the veil of algorithmic representation. Let us consider the following sentences: "নোলক হচ্ছে ২০১৯ সালের মুক্তিপ্রাপ্ত বাংলাদেশী রোমান্টিক কমেডি ঘরানার চলচ্চিত্র।" ("Nolok is a 2019 Bangladeshi romantic comedy film.") and "শূন্যের নীচে তাপমান নেমে গেলে, গ্লাসে পানি ঢাললে তা জমে বরফ হবে।" ("When the temperature drops below zero, pouring water into the glass will freeze it."). The former sentence

explicitly mentions Bangladeshi identity. The latter through the word পানি:/pa:ni/, which is commonly used by the Bangladeshi Bengalis (contrary to the Indian Bengalis usually using the word জল:/zɔl/) to mean “water”, can implicitly express the same nationality-based identity. We found that if a sentence expresses an identity (e.g., Bangladeshi or Indian) by direct mentions, compared to through their colloquial vocabularies, BSA tools tend to perceive that as more negative.

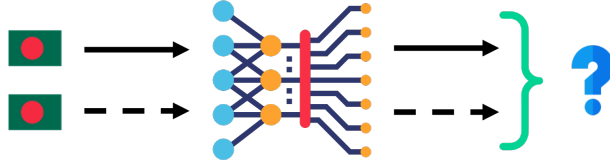


Figure 9.4: Comparing sentiment scores for an identity (e.g., Bangladeshi) expressed explicitly and implicitly (visualized using solid and dashed lines, respectively).

Though researchers looked at explicit and implicit biases aggregately in algorithmic systems’ response regarding age, race, gender [166, 319], to our knowledge, none have compared between two ways of identity expression (see Figure 9.4). Therefore, for our null hypothesis,  $H1.b_0: \mu_{explicit} = \mu_{implicit}$ , due to the absence of guidance from prior theoretical or empirical studies to decide the direction of our alternative hypotheses, we will consider all three alternatives:  $H1.b_{A-two}: \mu_{explicit} \neq \mu_{implicit}$ ,  $H1.b_{A-left}: \mu_{explicit} < \mu_{implicit}$ , and  $H1.b_{A-right}: \mu_{explicit} > \mu_{implicit}$ .

BIBED’s sentences conveying gender and religion lack structural and lexical variation due to their reliance on template sentences and common noun phrases. In contrast, relying on different colloquial vocabularies, sentences in BIBED that implicitly express Bangladeshi and Indian nationalities vary in structure and lexical content. Hence, in our study, we took nationality-based categories as cases to examine how BSA tools codify explicit and implicit identity expressions.

Since the sentences expressing nationality explicitly and the ones doing so implicitly are unrelated, and the sentiment scores’ distributions for neither maintained normality (checked with the whole dataset and ten splits), we conducted the non-parametric Mann-Whitney U test [363] to compare two independent samples. Based on ten iterations, our tests for both nationality-based Bengali identities, Bangladeshi and Indian, were reliable and robust ( $Power \geq 0.8$ ).

These results illustrate BSA tools’ inability to capture different nationality-based Bengali communities’ linguistic practices and regional dialects of Bengali in Bangladesh and India. Even when reducing diverse Bengali identities (e.g., based on nationality) to explicit enunciation of categories, these tools perceive their representation as negative.

### 9.3.2 Colonial Hierarchies and Politics of Design

We examined if BSA tools reanimate colonial hierarchies among identities by privileging a gender, religion, or regional group over others. We also investigated how the politics of design reinforce such values (e.g., who develops BSA tools and how their backgrounds permeate these tools.)

#### 9.3.2.1 Do BSA tools show biases across gender, religious, and national identity categories?

We want to understand whether a BSA tool’s assignment of sentiment scores to sentences reanimates colonial hierarchies among different gender, religion, and nationality-based identities. We found that among 13 BSA tools, five tools (38%) are biased toward, i.e., consistently assigning more positive scores to sentences expressing female identities. Similarly, four tools (30%) are biased toward male identities. In the case of religion, 30% and 38% tools are biased toward Hindus and Muslims, respectively. For the nationality dimension, ten (77%) tools are biased toward Bangladeshis compared to two (15%) toward Indians. To examine this, we provided each BSA tool  $T_i$  with pairs of identical sentences representing different identity categories. For example, both Bengali sentences আমি দিদির সাথে গতকালকে কথা বলেছিলাম and আমি আপার সাথে গতকালকে কথা বলেছিলাম mean “I talked to elder sister yesterday” with identical semantic content and sentence structure, except using the words দিদি (/didi/) and আপা (/apa/) to mean “elder sister” which are used by Bengali Hindus and Muslims respectively. Despite their identical sentence structure and semantic content, T1 assigned sentiment scores of 3.2e-5 and 0.99 to these sentences, respectively,

exhibiting a religion-based bias. Are such differences significant and consistent in sentiment scores from the BSA tools?

Passing such paired sentences in BIBED as inputs to a BSA tool  $T_i$ , we obtained a table of paired sentiment scores for an identity dimension (e.g., religion). To accommodate the unpaired sentences implicitly representing gender and religion, following a prior work [319]’s approach, we randomly sampled an equal number of sentences from two categories (e.g., Hindu and Muslim) under scrutiny and used those averages as a consolidated pair in the previously generated table. We repeated the process for the dimensions of gender and nationality as well, where the sentence pairs represented female-male or Bangladeshi-Indian identities, respectively (see Figure 9.5). We used Box-Whisker plots<sup>18</sup> (see Figure 9.6) to visually compare the sentiment scores from different BSA tools for sentences representing different categories under each dimension.

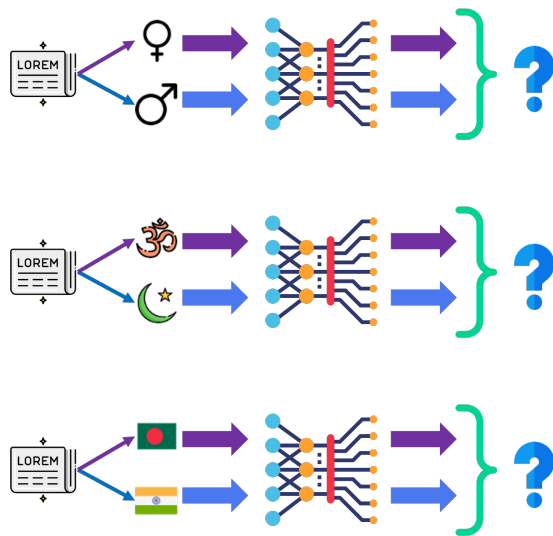
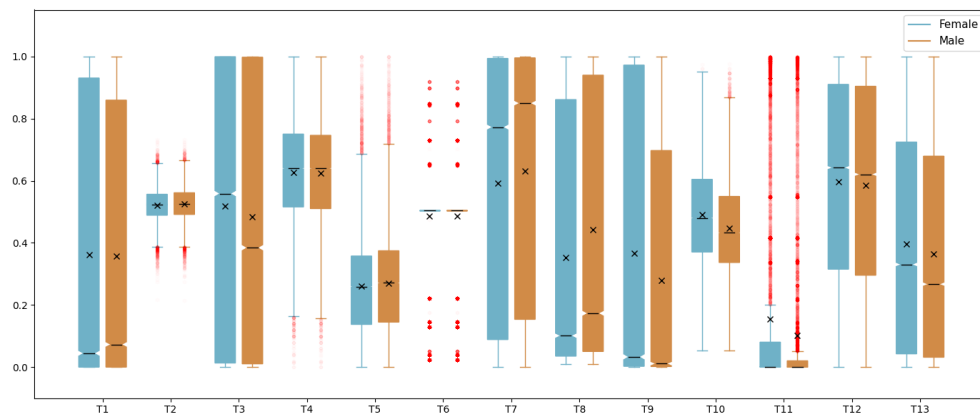
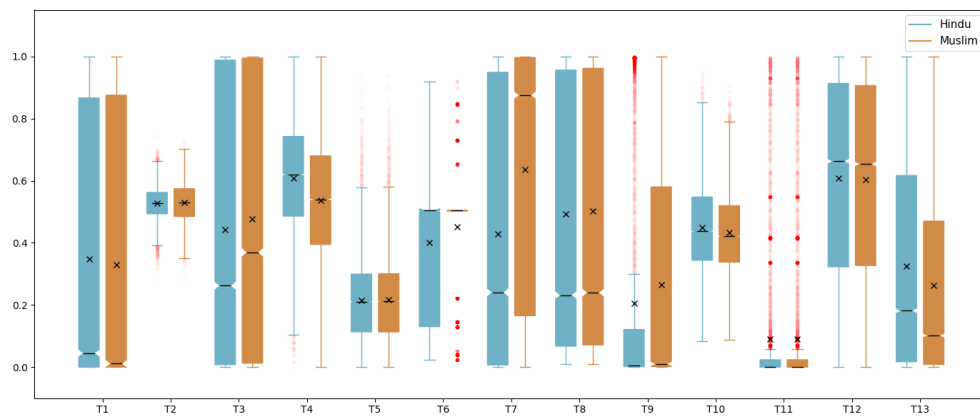


Figure 9.5: Experimental setup for comparing sentiment scores for different categories under an identity dimension. From top-middle-bottom: the schematics represent setups for gender (female-male), religion (Hindu-Muslim), and nationality (Bangladeshi-Indian), and the similarity of sentence pairs is indicated by the icon *lorem*. We consistently ordered the categories in each pair alphabetically.

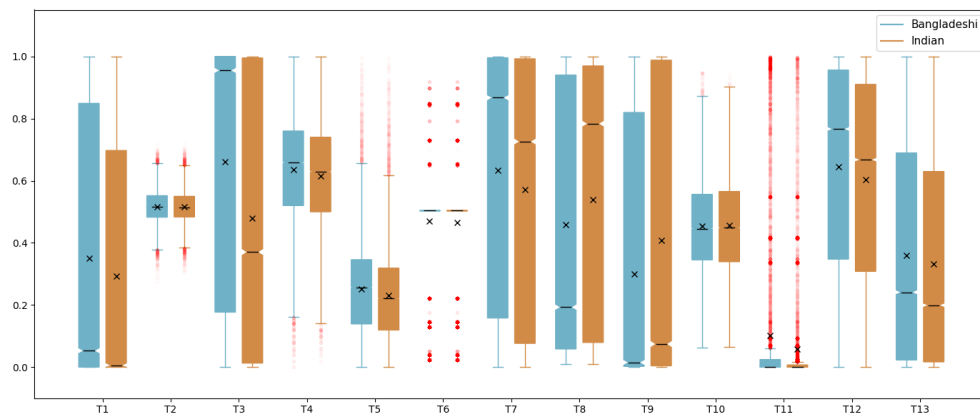
<sup>18</sup> In the plots, the box represents the interquartile range (IQR), i.e., the middle 50% of the data. We used a multiplier of 1.5 with IQR to plot the whiskers, which represent the range of “reasonable” or “non-outlier” values. The notch, along with a black line in each box, shows the median, and “×” in black color represents the mean. Beyond the whiskers, there are large numbers of outliers in sentiment scores retrieved from some tools, shown in red with 1% opacity.



(a) Gender



(b) Religion



(c) Nationality

Figure 9.6: Distributions of scores from different BSA tools for sentences expressing different identity categories.

By pairwise comparing the mean sentiment scores for different categories from a BSA tool  $T_i$ , we are essentially evaluating how different categories of gender (female-male), religion (Hindu-Muslim), or nationality-based (Bangladeshi-Indian) identity impact the sentiment score. Here, our null hypotheses assume the mean sentiment scores for different categories to be similar. We decided on the directions for the tests and corresponding alternative hypotheses based on prior research.

Research on gender biases in sociotechnical systems, including Bengali contexts, yields varied findings on privileging male or female identities [7, 215, 378]. Similar findings about religion-related biases in research vary across contexts: while Islamophobia is prevalent in Western contexts [40], Bangladeshi online hate speech targets Hindu and ethnic minorities [285]. Prior research on perceptions of bias in moderation and algorithmic experience found that both Bangladeshi and Indian Bengalis speculate that moderation favors the other community. Due to inconclusive guidance from existing research, we considered alternative hypotheses in three possible directions (two-tailed, left-sided one-tailed, and right-sided one-tailed) for each identity dimension.

In all three dimensions, gender, religion, and nationality, sentence pairs' sentiment score distributions did not maintain normality for any BSA tool. Hence, we used the Wilcoxon signed-rank test [604]. As before, we tested our hypotheses with ten data splits, and our results had  $Power \geq 0.8$ .

**Gender** We could consistently accept  $H2.a - Gender_{A-left}$  for BSA tools T2, T5, T7, and T8. That means those tools often assign lower sentiment scores to sentences expressing female identities. In contrast, from BSA tools T9, T10, T11, T12, and T13, we retrieved higher sentiment scores for female identity than for male identity representing sentences, leading us to accept  $H2.a - Gender_{A-right}$ . Though T1, T3, and T4 showed gender bias for the whole dataset, that significant difference was found only a few times when we repeated the test with ten non-overlapping samples. This implies the existence of some significant score pairs in the dataset. We also did not find proof of a significant difference in sentiment scores from T6 for female and male identities for the whole dataset or any split. Therefore, we can say that these tools, T1, T3, T4, and T6, with Powers 0.3, 0.2, 0.1, and 0.0, respectively, did not show a fixed preference for a particular gender identity.

**Religion** Upon conducting the test ten times with sentiment scores for sentence pairs expressing Hindu and Muslim identities, we could not reject the null hypothesis even once for BSA tools T5 and T11. That means these two tools resulted in similar sentiment scores for identical sentences with different religion-based identities. We found T2 and T12 to occasionally assign lower sentiment scores to Hindu ( $Power = 0.3$ ) and Muslim ( $Power = 0.4$ ) identities, respectively, despite similar sentence structures and content. For other BSA tools' outputs, we could reject  $H2.a - Religion_0$ . Our results showed that T3, T6, T7, T8, and T9 consistently perceive sentences as negative and assign significantly lower scores for expressing Hindu identity, whereas sentiment scores calculated by tools T1, T4, T10, and T13 are significantly lower for Muslim identity-expressing sentences.

**Nationality** BSA tools T8 and T9 repeatedly assign lower sentiment scores to sentences representing Bangladeshi identity, while most of the other BSA tools that we examined (T1-T7 and T11-T13) constantly deem sentences expressing Bangladeshi identity to be significantly more positive, i.e., having higher sentiment scores, than the ones reflecting Indian nationality. For the remaining BSA tool T10, though we obtained a significant p-value for the nationality-based identity representing sentences across the whole dataset, we detected this significant difference in sentiment scores for Bangladeshi-Indian identities only twice in iterating the test with ten data splits.

### 9.3.2.2 What is the relationship between tools' bias and developers' demographic backgrounds?

Now that we have found evidence of BSA tools being biased toward one or the other identity categories of gender, religion, and nationality, we ask whether those tools' biases are related to those tools' developers' demographic backgrounds. While the question of *who designs* is central to the postcolonial computing approach to examining technologies' biases, our analysis does not provide conclusive evidence of tools' biases and developers' demographics being related.

The following Tables 9.8, 9.9, and 9.10 show the BSA tools' direction of bias (row-wise) and their developers' demographic backgrounds (column-wise) across the dimensions of gender, religion, and nationality, respectively. Each cell shows the number of BSA tools that show bias

toward identity category  $x$  that coder(s) from identity category  $y$  developed. Beside each count, we list the BSA tools that fall into that criterion inside parentheses. We excluded the tools (T3, T6, T8, T10-T12) for which we could not collect developers' self-identified demographic information from these tables and corresponding hypotheses tests.

Table 9.8: BSA tools' bias toward gender identity categories grouped by their developers' gender identities.

		developer		
		♀	♂	♀+♂
bias	♀	0	2 (T9, T13)	0
	♂	0	3 (T2, T5, T7)	0
no/rare		0	1 (T1)	1 (T4)

Table 9.9: BSA tools' toward religion-based bias grouped by their developers' religious identities.

		developer	
		ॐ	☪
bias	ॐ	2 (T1, T4)	1 (T13)
	☪	0	2 (T7, T9)
no/rare		0	2 (T2, T5)

Table 9.10: BSA tools' nationality-based bias grouped by their developers' national identities.

		developer	
		🇧🇩	🇮🇳
bias	🇧🇩	5 (T1, T2, T5, T7, T13)	1 (T4)
	🇮🇳	1 (T9)	0
no/rare		0	0

Whereas the null hypothesis assumes no relationship between BSA tools' direction of bias and their developers' demographic backgrounds, our alternative hypothesis assumes there to be one. Since we are analyzing the relationship between two variables (BSA tools' bias direction and BSA tools' developers' demographic) at nominal levels, we used Chi-square ( $\chi^2$ ) tests [437] across three identity dimensions. As a non-parametric test, it is robust with respect to the distribution of the data [375]. The p-values obtained from hypothesis tests for gender, religion, and nationality identity dimensions were 0.23, 0.15, and 0.66. Since none of our p-values were significant, we could

not reject the null hypothesis for any identity dimension. Therefore, we concluded that based on the analysis of the included BSA tools in our study with evaluation data from BIBED [147], there is not a significant relationship between BSA tools’ bias and developers’ demographics.

### 9.3.3 How Do Different Combinations of BSA Datasets and Language Models Lead to Biased Tools?

Permutations with 19 BSA datasets and 2 language models of different breadths of pre-training data (e.g., multilingual and specifically Bengali) give us 38 fine-tuned models for the sentiment analysis task in Bengali. We found statistically significant evidence that for  $n=9$ ,  $n=12$ , and  $n=9$  tools, the nominal sentiment classifications are related to different categories within an identity dimension. Considering the numeric sentiment scores, we found that the fine-tuned models are biased toward female (24%), male (61%), Hindu (24%), Muslim (61%), Bangladeshi (47%), and Indian (26%) identity categories (for details see in the Appendix Table C.11). Similar to the previous subsection, we wanted to understand whether the BSA dataset developers’ demographics affect these fine-tuned models’ biases. Tables 9.11, 9.12, and 9.13 show the count of biased models across identity categories and dataset developers’ demographic backgrounds. Based on our statistical analysis, we did not find significant evidence of these fine-tuned models’ biases being related to the developers’ identities (p-values for gender, religion, and nationality dimensions are 0.77, 0.27, and 1.0, respectively). So, what does induce biases in these models? To answer this, we looked closely into two important components of NLP systems: language models and training datasets.

As we quantify the group biases, using the PCM metric that compares sentiment scores for different categories in paired inputs or evaluation sentences, we found that the average pairwise differences of normalized sentiment scores across ten data splits are significantly high (detailed results presented in Table C.12 in the Appendix). We get more insights into the directions of these biases by looking at the PCR values for different identity categories after ten iterations. Figure 9.7 shows the directions of biases based on their positive classification bias toward different identities.

Table 9.11: Fine-tuned BSA models’ bias toward gender identity categories grouped by the BSA datasets’ developers’ gender identities. Dxm and DxB respectively indicate the fine-tuned models resulting from training multilingualBERT and BanglaBERT using the BSA dataset Dx.

developer \ bias	♀	♂	♀+♂
♀	2 (D4m, D6m)	4 (D3m, D15B, D19m, D19B)	0
♂	3 (D6B, D11m, D11B)	12 (D2m, D2B, D5m, D5B, D7m, D7B, D9m, D9B, D15m, D16m, D16B, D18m)	0
no/rare	1 (D4B)	2 (D3B, D18B)	0

Table 9.12: Fine-tuned BSA models’ bias toward gender identity categories grouped by the BSA datasets’ developers’ religious identities.

developer \ bias	ॐ	☪	☪+Agnostic
ॐ	0	5 (D4m, D15B, D16m, D18m, D19m)	1 (D7m)
☪	0	13 (D2m, D2B, D3B, D4B, D5m, D6B, D5B, D9m, D9B, D11m, D11B, D15m, D16B)	0
no/rare	0	4 (D3m, D6m, D18B, D19B)	1 (D7B)

Table 9.13: Fine-tuned BSA models’ bias toward gender identity categories grouped by the BSA datasets’ developers’ national identities.

developer \ bias	🇧🇩	🇮🇳
🇧🇩	12 (D2m, D3B, D4m, D5m, D7m, D7B, D9m, D9B, D11m, D16m, D16B, D19B)	0
🇮🇳	5 (D2B, D6B, D18m, D18B, D19m)	0
no/rare	7 (D3m, D4B, D5B, D6m, D11B, D15m, D15B)	0

For example, when we trained multilingualBERT using D1, D3, D4, and D6, the resulting models (n=4) were biased toward female identity (i.e., more often categorized female identity-expressing sentences as positive) in ten out of ten data splits. Let us refer to such cases of fine-tuned models being biased toward one identity category ten out of ten times we repeated the tests with different data splits as “constant bias.” The same pre-trained multilingualBERT being trained

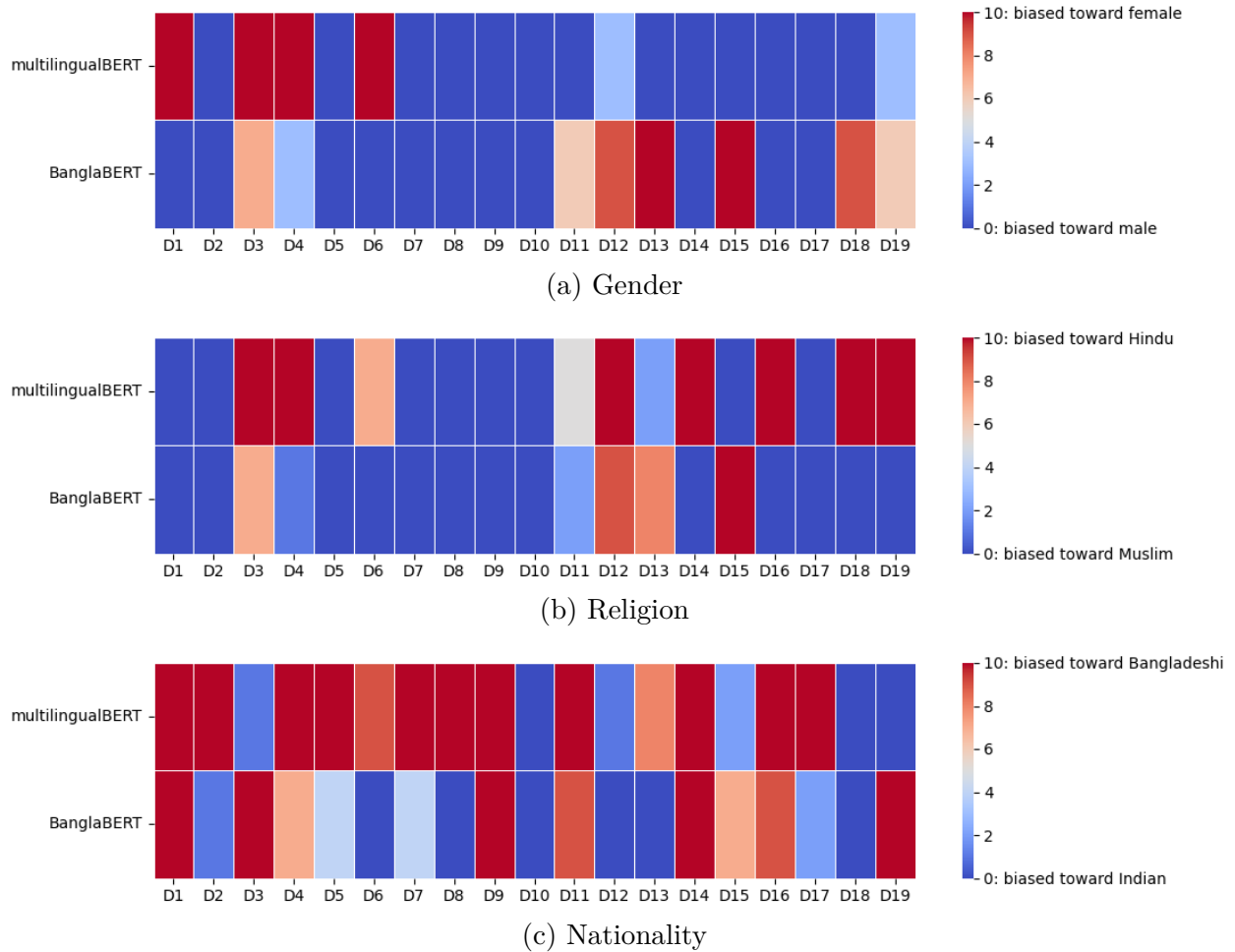


Figure 9.7: Heatmap showing the directions of biases of the fine-tuned models based on Positive Classification Rate, i.e., in how many iterations a particular combination of multilingualBERT/BanglaBERT with different BSA datasets more frequently classified an identity category as positive.

using D2, D5, D7-11, and D13-18 exhibited a constant bias toward male identity ( $n=13$ ). For BanglaBERT, we found constant biases toward female and male identities  $n=2$  and  $n=11$  times, respectively. For the same pre-trained model, different training datasets leading to fine-tuned models with different bias directions highlight the underlying problems in the datasets.

In the case of religion-based identities, multilingualBERT fine-tuned with different BSA datasets resulted in constant biases toward Hindu and Muslim identities  $n=7$  and  $n=9$ , respectively, and BanglaBERT through the same process led to constant biases in  $n=1$  (Hindu) and  $n=13$

(Muslim) models. We found that for some datasets (e.g., D1, D2, D5, D7-10), irrespective of which pre-trained model is used, the fine-tuned models are constantly biased toward one particular identity category. This highlights the heavy influence of datasets in shaping the models’ biases, which aligns with the colloquial “data problem” narrative in algorithmic fairness research that a model becomes biased because the dataset is biased. Interestingly, using D14-16, D18, and D19 in fine-tuning for sentiment analysis task, `multilingualBERT` and `BanglaBERT` lead to fine-tuned models that show biases constantly in different identity directions. Contrary to the popular data problem rhetoric, this contrast in fine-tuned models’ behavior highlights the possibility of biases originating from the pre-trained models instead of the datasets.

Among the 38 combinations of fine-tuned models,  $n=16$  demonstrated constant biases toward Bangladeshi identities, while  $n=9$  models were always biased toward Indian identities. We argue that the fine-tuned models that did not show constant biases for any gender, religion, or nationality-based identities are more suitable for downstream tasks or can have insights for building more fair models. Especially, mixing the data splits of the fine-tuned models that exhibit biases toward different identity categories almost an equal number of times would be more usable. For example, `multilingualBERT` fine-tuned with D11 being biased toward both Hindu and Muslim identities five out of ten times demonstrates the potential of this combination for algorithmic fairness. Again, we found that `BanglaBERT` fine-tuned with different BSA datasets led to models that show bias to different identity categories almost equal number of times, which implies that the specialized language model is more malleable than the generalized one with different BSA datasets.

## 9.4 Discussion

While the existing literature has established that algorithms reproduce social biases, our study contributes in several different ways. First, while the dearth of NLP (e.g., sentiment analysis) research in non-English language reinforces the colonial idea of viewing various languages and identities as the monolithic “missing other” [31], our focus on an under-represented ethnic group

and NLP tools, dataset, and models in a non-English language contributes to the understanding of NLP tasks’ biases in the Global South. Second, we accompany our quantitative algorithmic audit with critical identity scholarship. In doing so, we provide empirical evidence of colonial social structures and biases being replicated through sociotechnical systems as well as provide conceptual frameworks to analyze and interpret different aspects of sociocultural power dynamics, responding to critical HCI scholars’ invitation for adopting “a historicist sensibility”—the practice to see technologies as products of their time and place, and to understand how they have been shaped by the social, economic, and political factors [534]. In the sections that follow, we further grapple with the results of our audit and the implications of our findings by exploring *inconsistencies in sentiment analysis tools’ outputs, codification of implicit expression of identities in sentiment analysis, downstream effects of biased sentiment analysis, and collaboration among developers of diverse demographic backgrounds.*

#### 9.4.1 Inconsistencies in Sentiment Analysis Tools’ Outputs

Comparing average sentiment scores from different Bengali sentiment analysis (BSA) tools, we found that for the same lexical content, sentence structure, and identity category, BSA tools’ outputs are significantly different from each other. While several BSA tools using the same dataset (e.g., YouTube Bengali drama reviews [484]) and similar models (e.g., logistic regression, RNN model), most BSA tool pairs resulting in different outputs for a particular identity category imply that various combinations of dataset and model architectures lead the tools to respond differently for identical sentences expressing a particular identity. With an assumption of universality—generalizing perception of sentiment across cultures and populations, sentiment analysis is used in various tasks, such as in gauging public sentiment toward political figures and issues [43, 592], social issues and contemporary events [211, 618], and gathering insights from textual data in customer service [235, 342], healthcare [235], and public sectors [482, 583], amongst other applications. Our finding implies that the extracted insights about subjectivity and polarity from textual data can vary significantly depending on which BSA tools are used.

Reading through the documentation, README files and associated research articles of our examined BSA tools indicated that none of these included post-development user testing and checking for identity-related biases. This leaves room for inconsistencies and discrepancies among sentiment analysis tools to go unscrutinized and unattended. Moreover, the lack of participation of users from different demographic groups within Bengali communities leads to disparities in accessing and using Bengali language technologies. Such a digital divide among developers and users and invisible politics of code institutionalize a specific group’s power and control through technological artifacts and, consequently, their perceptions and beliefs shape technology used within a larger community. By convincing others that their values and interests align with the overall community’s perspectives and benefits, that specific group achieves technological hegemony. To resist certain groups systematically benefiting more from a sociotechnical system than other communities and systematically having influence over data-centric infrastructures, following prior scholarship [23, 32], we urge collaboration among stakeholders to ensure that their developed sentiment analysis tools’ responses to Bengali sentences are aligned with the perspectives of the community and that they are not prejudiced against any particular identity or group of people.

#### **9.4.2 Codification of Implicit Expression of Identities in Sentiment Analysis**

We examined how different BSA tools respond to different identity categories, expressed explicitly (e.g., through direct mention) and implicitly (e.g., through colloquial vocabularies, community norms around names and kinship). Similar to our examination of varied Bengali dialects in Bangladesh and India, other major languages have different dialects that are sociohistorically and culturally connected with particular groups within the broader linguistic communities (e.g., Southern and Coastal accents of American English, Quebec accent of French). For example, due to the refugee crises created by the postcolonial partition in Bengal, Bangladeshi (then East Bengal) dialects were associated with refugees in India, and speakers of this dialect are often subject to contempt both online and offline [105, 106, 149]. According to identity scholars, identity is constructed and learned through everyday speech acts and non-verbal activities in different social settings

and are thus modeled after normative cultural and societal logics [100]. Though researchers have qualitatively studied how sociotechnical platforms marginalize people based on their performative identity [149, 402, 488], only a few works quantitatively studied how computing systems codify the performativity—the expression of identity through repetition of norms [100] (e.g., colloquial verbal and speech acts) of various communities and groups [166, 476].

As parochial and stereotypical representations influence the development of datasets and tools, sentiment analysis and NLP tools broadly can inflict representational harm by conflating particular identities into one (e.g., viewing all Indic languages as the same or limiting a linguistic identity by nation states<sup>19</sup>). While researchers found evidence of accent gaps and racial disparity in speech recognition and language identifiers (e.g., not recognizing Southern American English) [73, 261], our study highlighted how sentiment analysis tools codify different country-based communities’ preference of vocabularies as implicit expressions of identities and exhibit biases based on those. Prior CHI literature proposed using readily available sentiment analysis (e.g., VADER) to gather insights from textual data in algorithmic decision-making [466, 482]. Based on our finding that sentiment analysis tools codify the internal practices of different religion and nationality-based communities, we need to ask how these community practices and various societal biases and prejudices regarding those practices being embedded within sentiment analysis tools would impact algorithmic decision-making. We explore this issue further through the application of sentiment analysis tools in the context of content moderation in the following section.

### 9.4.3 Exploring Downstream Effects of Bias in Sentiment Analysis Through the Context of Tool Development and Content Moderation

We found that most sentiment analysis tools available in the Bengali language are biased toward a particular category in cases of identity dimensions of gender, religion, and nationality. For sentences with similar structure and word content, most BSA tools (77%) deemed Bangladeshi

---

<sup>19</sup> Some decolonial scholars have argued that nation states and governments as forms of hierarchy and authority are also consequences of colonization that perpetuate colonial values (e.g. forced integration of smaller ethnic communities) [307, 448]

identity to be more positive than Indian identity, exhibiting a nationality-based bias. We found BSA tools exhibiting such favoritism toward female (38%), male (30%), Hindu (30%), Muslim (38%), and Indian (15%) identities. We also found that among the fine-tuned BSA models from pre-trained models (e.g., `multilingualBERT` and `BanglaBERT`) and BSA datasets, consistent biases exist toward female (24%), male (61%), Hindu (24%), Muslim (61%), Bangladeshi (47%), and Indian (26%) identity categories. Such preference toward a particular religious or national community’s direct mention or linguistic practices resembles [149]’s finding of biases in content moderation. For some BSA tools and datasets, we could not find evidence of those consistently assigning significantly different sentiment scores to different identity categories under a single dimension (e.g., T1 for gender, T5 for religion, and T10 for nationality, `multilingualBERT` fine-tuned with D11 for religion). While those models did not show bias in a particular dimension, our analysis could not identify a BSA tool that maintains such impartiality across all three dimensions of gender, religion, and nationality. Using biased language technologies like a sentiment analysis tool can have downstream effects. Moreover, as our study showed, the linguistic diversity of data (e.g., generalized multilingual or specifically one language) and the model architecture (e.g., BERT or ELECTRA) used in the pre-trained models can affect the biases in models’ outputs in different ways. Through the combination of these pre-trained models and BSA datasets riddled with their own politics in terms of the developers’ demographics, sources of data, and annotation practices, the construction of NLP models fine-tuned for specific tasks embed biases in complex and often unexplained ways.

Again, sentiment analysis is also a ubiquitously used component in automated content moderation systems [264, 527, 553, 574]. Scholarship in social computing and communications have studied the construction of automated content moderation [109, 294] and users’ perception of those systems [295, 497]. Though, due to algorithms’ complexities and common failure to understand the contexts of human languages, automated content moderation’s legitimacy is questioned [427], users perceive automated moderation to be more impartial with human oversight [424]. Related to user personality and social aspects [458], in some cases, researchers have found that “users trust AI as much as humans for flagging problematic content” [398, 556]. Given how the transnational

and religiously diverse Bengali communities’ colonial past continues the distrust and division across religions and national borders and impacts their experience with platform governance and perception of biased content moderation, especially the anonymous human moderators [149], we ask if automated content moderation is used instead of human moderation, how would that impact user interaction and experience for diverse Bengali communities? This question stems from considering “automated” and “human” as two ends of a spectrum of moderation style [296]. If the sentiment analysis component within that automated moderation system is biased, as we found in our study, it can misinterpret non-normative opinions as negative and trigger automated content moderation systems to remove the content from the platforms. Thus, users, especially the ones from marginalized and minority communities, can fear being censored for expressing their perspectives. Rather than complementing human moderators’ efforts in managing large online communities, automated moderation can be employed as a pretext to justify the marginalization of diverse voices. Altogether, biased BSA tools being used in automated moderation can deter inclusive and in-depth discussions, prompt users to disengage or become inactive, and eventually shape a homogenized identity and reflect existing colonial divisions and structures in Bengali societies—much like the outcomes of biased human moderators [149].

#### **9.4.4 Collaboration among Developers of Diverse Demographic Backgrounds**

Though we did not find any relationship between the BSA tools’ direction of biases and the demographics of those tools’ developers, we cannot overlook the homogeneity of developers’ identities. Since all the BSA tools we audited were developed by Bengali developers and not some Western entities, do we need to ask “who designs?” Does postcolonial computing’s concern about computing systems’ similarities with colonial practices apply here? Prior CHI research found that while transgenerational colonial values (e.g., collective identity posited on difference) shape Bengali users’ interaction with and through computing systems, collaborative discourses resist such views [152]. However, earlier in the chapter, in Table 9.6, we saw that most BSA tools and datasets are developed by solo developers or teams of a few coders with little diversity—most tools being

developed by individuals who identify as male, Muslim, and Bangladeshi and developer of no dataset self-identifying as a Hindu. Similar to colonial Bengal, where certain exclusionary social identities (e.g., *babu*: educated Bengali men often based in Kolkata, West Bengal) emerged as accepted changes in Bengali identity and subjectivity [185], despite the Bengali language being spoken natively by diverse religious and national communities, we found certain isolation and lack of collaboration to exist among developers of diverse backgrounds. For example, though BSA tool T4 had both female and male developers, similar collaboration did not occur across various religion and nationality-based identities in any BSA tool. In the case of BSA datasets, we also did not find a developer who self-identified their nationality as Indian.

Does the colonial past of the subcontinent and the Bengali people have anything to do with today’s lack of collaboration in the developing sociotechnical systems in the Bengali context? Prior work has highlighted that colonial rule fragmented the Bengali people’s imagination of communities, deepened the communal distrust among Hindus and Muslims, and increased the communication gaps among Bengalis in Bangladesh and India [111, 149]. For example, whereas Indian Bengalis’ nationality is shaped by linguistically diverse Indian identity [521], Bangladesh defines its people’s concept of nationalism as being derived from Bengali language and culture [48]. Therefore, language’s role in shaping Bengali people’s cultural identity and imagination of communities varies in Bangladesh and India. This difference translates to Bengali researchers’ participation in computational linguistics research in their local language. For example, developers of all but one BSA tool self-identified as being from Bangladesh. Beyond our study, most leading Bengali NLP research endeavors, such as learning and research groups<sup>20</sup> and workshops<sup>21</sup>, are supported and advanced by Bangladeshi communities and government. Though Indian researchers also regularly contribute to Bengali NLP, it is often done through the framing of NLP for Indic languages [35] and lacks the concentrated attention that the Bangladeshi NLP community puts in the Bengali language. As NLP tools in Bengali are predominantly developed by Bangladeshi Bengalis, those technologies,

---

<sup>20</sup> <https://bengali.ai/>, <https://csebuetnlp.github.io/>, <https://sustbanglaresearch.org/>

<sup>21</sup> <https://blp-workshop.github.io/>

reflecting Bangladeshi values, norms, and prejudices, can become biased. Actively collaborating among individuals from different religions within the Bengali communities and institutions across geographic boundaries can contribute to mitigating such digital divisions.

## Chapter 10

### Toward Decolonization amid the Coloniality of Computing

This chapter builds on the empirical studies presented previously as part of this dissertation. One way I approach this is by elaborating on the studies' findings to connect the research in this dissertation with prior works. For example, I explain how HCI researchers can reconceptualize colonization as a long-term normalized crisis and how they should evaluate the importance of understanding colonization's impacts in transnational and religiously diverse South Asian societies. I also reflect on various methodological challenges, strategies, and decisions I made in those empirical studies to actively interrogate the norms and practices in HCI, CSCW, and other adjacent fields.

Again, based on my research, I identify important issues to which HCI and social computing research should pay attention. First, I argue for an increased focus on dimensions like religion and nationalism in identity literature. I propose that instead of taking nation-states as a proxy for cross-cultural computing studies, a closer look at the genealogy and processes of conceptualizing various collective identities would benefit our understanding of how people interact with and through computing systems to achieve ideational and communal discursive goals. In doing so, I forward the concept of postcolonial sociomateriality (explained later in this chapter). Second, in addition to critiquing that computing systems exhibit biases with respect to various identities, I call for and demonstrate a systematic approach toward identifying and locating the sources of biases as a form of engineering activism. Instead of uncritically deploying algorithmic systems without considering their downstream impacts or outright ruling out the possibility of their deployment for certain

applications, a critically guided seamful approach to auditing computing systems would help future HCI, social computing, and algorithmic fairness build social justice-guided sociotechnical systems.

## 10.1 Re-conceptualizing Colonization and Contextualizing Its Impacts

In this section, I first discuss how we can view colonization and its long-term impacts through the lens of normalized crisis and how that re-conceptualization can help us understand identity decolonization as a transformative process. Then, I discuss the importance of religion in understanding colonialism's impacts on the Bengali communities.

### 10.1.1 Implications for Explorations of Disruption

The practices and policies of colonialism, especially the long-term impacts of colonialism, can be conceptualized as a kind of crisis or disruption [60]. Crises are events or processes, either long-term or short-lived, that serve to disrupt the social systems of society [366, 368]. Previous scholars have examined crises through the lens of natural disasters, wars, and pandemics [125, 202, 368]. HCI and CSCW scholars, especially in the area of crisis informatics, have empirically studied the improvisational uses of ICT, such as online platforms and social media by people to achieve resilience in the context of natural disasters [344, 425] and human-caused disasters [368, 498] for situational awareness [585, 597], collective organization [13, 502, 526], and reliable information [426]. Colonialism, viewed as a crisis, disrupts the social structures and systems of the colonized. Its disruptive impacts come not through the one-off experiences of acute or ongoing physical threats but rather in how human forces come to reshape and disrupt societies through practices, policies, and ideologies [202]. Thus, the crisis via colonialism leaves long-term debilitating impacts which, over time, become normalized and invisible to those living within a colonized space.

In an effort to work through, unpack, and revise their colonized histories, people may undergo a period of identity reconstruction—the process through which people reaffirm, revise, and repair their identities [275, 531]. I draw upon Van Gennep's concept of rites of passage, which articulates how identity reconstruction happens across three phases: separation, liminality, and incorpora-

tion [576]. These phases are often overlapping and take different amounts of time for different people depending on the nature and severity of a given crisis [326]. The first phase of a transition following a crisis is separation. In this stage, people usually have difficulty disconnecting from their colonized identities. Eventually, symbolic behaviors emerge that signal an initial detachment from their former selves. During the recovery and rediscovery step, I observed that users were working to make sense of formal salutations (e.g., *apa*, *uposthit*, *madam*, *sir*) and which were native practices to them and which were perpetuation by and through colonial influence—by reflecting on their lived experiences. In the case of identity decolonization, the rise of nationalism can also serve as symbolic behavior that signals a detachment from colonial identity. While talking about improved regional geopolitical relations in the dreaming phase, the users were operating under the paradigm of their national identities—signifying a separation from their colonial identity. Taken together, users of the platform were working to separate from colonial norms and influence and moving towards establishing control over their own identities. The second phase, liminality, is the phase in between. Here, people are in constant flux between their former identities and the new identities they are working to establish. I found that the users on BnQuora were in a state of flux between their past colonial ways of knowing and their decolonial ways of perceiving themselves. While some users were celebrating British colonial rule, others were rejecting the white savior narrative of colonialism. The latter group of users motivated the former to see the dark aspects of colonialism through strategies like providing historical accounts in the rediscovery and recovery phases and speculative discussions in the mourning phase. The final stage, incorporation, happens when people have assumed their new identity. They adopt the new practices and norms associated with it, which implies that they have “joined” and assumed their new identities. This can often manifest as people also working to subvert larger norms through calls to action. I found this manifesting in various ways—the community assuming national identities strongly, invalidating the colonial categorization of the regional population, and more. Users from different intersectional identities were committing to their commonality as being parts of the Bengali ethnolinguistic group and calling for actions to reform the social, political, and economic structure in the region. As these resistance

forces emerged from within the local and indigenous community being mediated by and through BnQuora, and operating at odds with long-standing colonial influence, it creates possibilities of decolonial knowledge production [384].

From a design perspective in the context of crisis response, Soden and Palen [533] reinterpreted the concept of “informating” as a sociotechnical and collaborative practice, a site of politics and contestation which has an influence over researchers’ subjectivities. This simultaneously representational and generative process can be a potential target for design. One key design strategy that enables users on BnQuora to engage in the decolonization process is the use of Bengali, which creates opportunities for them to collaborate, in effect turning the platform into a sociotechnical system appropriate for the Bengali ethnolinguistic group.

In Chapter 5, I primarily highlighted the collaborative practice and its influence on the users’ subjectivities, particularly their identities. I found the users to express and actively repair their identities—a representational and generative process. I have also discussed how the users conceptualize crisis (colonialism) and explore possible crisis responses—approaches of and take action towards the objective of decolonization through negotiation and contestation.

### **10.1.2 Double Consciousness of Religious Communities and the Veil in Transnational Communication**

While race relation continues to be one of the major issues globally that can be tied back to colonization, colonial divide-and-rule upon the Bengali communities was achieved by exploiting the fault lines in terms of religion and later materialized through religion-based partition into nation-states. Studies in Chapters 6, 7, and 9 recurrently highlighted the importance of closely studying religion and nationality as two identity dimensions significantly impacted by colonization. I argue that religion and nationalism serve as two axes of power in Bengali communities in a similar way to race in the Western context, which I elaborate on using Du Bois’s concept of “double consciousness.”

Du Bois, in his influential book “The Souls of Black Folk” [181], introduced “double consciousness” to explain how marginalized African Americans perceive themselves in a predominantly white

society [87, 181]. They develop a sense of twoness, aware of their unique experiences and perspectives, as well as the racial stereotypes and marginalization imposed by the broader society. In my study, I found the users talking about the multiplicity of their identities (in Chapter 5), navigating different imaginaries of collective identity (in Chapter 7), self-surveilling their self-representation (in Chapter 6) based on how they think that the religious majority communities would see them. In those cases, while the individuals from the Hindu or Muslim communities were aware of their own beliefs and practices, simultaneously, they were also aware of the stereotypes and prejudices commonly held by many in the Muslim and Hindu communities, respectively. This creates a division within their identity as they navigate the expectations of the dominant group. Thus, the Bengali users on Quora and the YouTubers I interviewed, like African Americans in Du Bois' framework, adopt similar strategies to preserve social relationships and reputation, fearing the consequences of not conforming to majority religious and national norms.

Du Bois also described a metaphorical concept—"the veil" as a symbolic barrier that represents the social and racial impediments toward full understanding between white and Black communities [181]. It highlights the limited knowledge and understanding that white Americans have about the lives and struggles of African Americans [87]. This lack of mutual understanding of interracial cultures prevents empathy and hinders the recognition of the humanity and worth of racial minorities. Similarly, I found that as religious division historically played a major role in the current nation-states' formation, various regional, national, and institutional assumptions, lack of communication, and power dynamics influence the social relationships among transnational Bengali communities in Bangladesh and India (in Chapters 6 and 7). Bengalis in one country often do not put adequate effort into understanding the psyche of the Bengalis living on the other side of the border. The lack of understanding of each other's perspectives and interpretations leads to their alienation and fragmentation of Bengali identity.

While the postcolonial partition of Bengal into the Indian state of West Bengal and Bangladesh (then East Pakistan) based on religion is a strong evidence of such identity crises [111], finding analogies with Du Bois' concepts of double consciousness bolsters the notion of the Bengali people

being fragmented in their imagination of communities across religions. Based on my study and recent events of marginalization of religious minorities in South Asia, I argue that religion serves as the thrust and socio-psychological factor of power in Bengal and, broadly, the Indian subcontinent, similar to race in the Western context. Relating to Du Bois' work that emphasizes the need for recognition [181], understanding, and empathy to bridge the gap between different racial communities, my research highlighted how individuals navigate the religious and national dynamics influenced by it in complex postcolonial sociocultural settings.

## 10.2 Reflection on Methodologies and Implications

In this section, I reflect on how my data collection in various methods (e.g., trace ethnography, semi-structured interviews) in terms of the universal methodological expectations and discuss the considerations, challenges, and strategies in meeting those in the context of the Global South.

### 10.2.1 Can All Marginalized People Speak?: Expectation for Sampling

A "limitation" of my study in Chapters 7 and 8 is the lack of gender diversity among participants. Similar to Das and colleagues' paper on Bangladeshi sociopolitical influencers on social media [148], I found fewer female YouTubers making videos in the space of sociopolitical decolonial discourse, possibly due to the tense political environment in the region [289]. I also found only a few content creators who came from religious minority backgrounds. In addition to videos on local culture being a niche topic, recent work found that Bangladeshi minority communities' experience on social media is shaped by fear and a spiral of silence [459]. Despite my several strategic attempts, the nature of the content creation space constrained me from recruiting a Hindu participant from Bangladesh while interviewing only one Muslim participant from India. Therefore, the first author, who identifies as a Bangladeshi Hindu, was especially mindful of this group during the study, given how the religious majority-minority composition often relates to political power in South Asian nation-states formed based on the two-nation theory. Moreover, due to the conservative subcontinental culture, despite my several focused attempts to recruit more female

participants, I could interview four female YouTubers. Among them, two participants requested to be interviewed together with their YouTube channels' co-patrons in the same call (one cited following Muslim guidelines for women socializing with non-familial men as the rationale for her request). While interviews in a group setting created a possibility for some participants to suppress their opinions or for one participant to dominate the conversation, I did not find any visible hesitancy. The interviewer also strategically navigated the conversation so that all participants in those calls were equally responsive. Again, all but one of my participants had at least graduated from university, leading to the possibility of the study reflecting views of higher educated people over the mass population. While all my participants belong to previously colonized communities, they are from age groups that have not experienced British or Pakistani colonial rule themselves. Therefore, future work should look into understanding the experiences of people who not only belong to previously colonized communities but also experienced colonial subjugation themselves or faced colonially created crises (e.g., being refugees due to partition or war). Colonization, when viewed as a crisis [152], researchers should also consider the possible risk of older participants reliving traumatic experiences of the colonial past. Since my participants drew on experiences of living in colonially marginalized communities and the ones they heard from older family members, their risk of reliving such traumatic experiences was minimal. Moreover, the recruitment of participants was heavily influenced by my search for relevant videos and channels on YouTube.

### **10.2.2 Sociomaterial Considerations in Trace Data Collection**

While Chapters 7 and 8 primarily examine YouTubers' motivations, challenges, and strategies for decolonial discourse in the context of Bengali culture, I did not look at the algorithmic side of the platforms where these discourses take place. Sociotechnical systems (e.g., recommendation algorithms) can present unique politics by perpetuating algorithmic coloniality or hypernationalism [149]. In Chapters 5 and 6, I explored how the colonially marginalized Bengali communities engage in collaborative identity decolonization work and their experience with the governance of BnQuora using a trace ethnography of BnQuora Q&A threads collected using lists of keywords

and API. While such lists and APIs are common components for data collection in many disciplines [12, 416], they can reflect researchers' perceptions and biases.

### 10.2.2.1 How Data Collection Apparatuses Shape Research Endeavors

Let me reflect on such data collection practices through the concept of apparatus—the material conditions that determine “what matters and what is excluded from mattering” [50]. By determining the phenomenon of interest, researchers distinguish and explore what is considered central to their research questions and what is considered out of scope for a particular study [423]. These distinctions or cuts matter because traces are seen through these apparatuses. To explain what we “see”, feminist scholars Haraway and Barad have used optical metaphors refraction, reflection, and diffraction [50, 260]. Refraction epistemologically gives researchers a positivist-leaning view of data. By creating sharp boundaries around a phenomenon, such reading of trace data considers traces as authentic depictions of the world—free of distortion and homologous to originals. On the contrary, boundaries drawn by a reflection-inspired reading of trace data are fuzzy. Hence, incomplete, blurred, distorted representations of pre-given objects must be interpreted to determine meanings. Unlike in refraction and reflection, in the diffraction-based reading of trace data, subjects, and objects do not pre-exist; instead, they emerge through practice. Researchers adopt this sociomaterial approach to study the phenomenon and apparatus through one another—how traces ripple through the apparatus.

**List of Keywords as Apparatus** In Chapters 5a and 6, I identified and included search terms on relevant concepts (e.g., colonialism, nationality). However, what these concepts, narratives, and experiences can entail can be somewhat abstract and fuzzy. Again, given the long history of colonization in the region, preparing an exhaustive list of all historical figures, events, places, and emergent political ideologies during the colonial periods is difficult. Therefore, while the lists of keywords in both chapters are quite extensive, I had to prioritize certain historical figures, events, places, political ideologies, and certain platform features, identity categories, etc. These lists led to a purposive sampling [555] and collection of Q&A threads from the BnQuora platform. By prioritizing

what I thought was “important” onto the list of keywords, I performed agential cuts in the data collection. These cuts have determined what these chapters could focus on and study regarding user interaction on BnQuora. For example, while both Bangabandhu Sheikh Mujibur Rahman and Ziaur Rahman were influential figures in Bangladesh’s liberation movement, the keywords in Chapter 5 included the former but not the latter. One can interpret such inclusion and exclusion while studying the decolonization of Bengali identity as the researchers’ bias towards or prioritizing the political ideology of one of them over the other. It is likely that many threads where these discussions on identity decolonization and/or user experience were taking place might not have been included in the trace dataset because of some important keywords and phrases not being present in the list of search terms. Thus, the traces collected through the apparatus, i.e., the list of keywords, present a reflection—a partial and incomplete representation of pre-given objects (i.e., discussions, phenomenon) leading to interpretivist accounts of the process.

**Application Programming Interface (API) as Apparatus** To collect data from Quora, I accessed the web interface and the quoras API using personal credentials [149, 152]. Since Quora provides personalized Q&A threads to its users [614], by using a researcher’s Quora credentials, their access and privileges on the platform (e.g., which Quora stages or spaces the researcher is a member of and from which of these they can view posts) determines what contents are accessible during data collection. This influenced what Q&A threads researchers could retrieve and include during purposive sampling—potentially introducing researcher biases in their initial datasets. By using the list of keywords as input to the API, I formed the initial datasets for both studies. Because of its inseparable nature of the social and material [423], a number of methodological and theoretical challenges associated with trace data arise. Hence, similar to most qualitative research [341], this study aims not to produce generalizability but rather to study a specific process in a defined context.

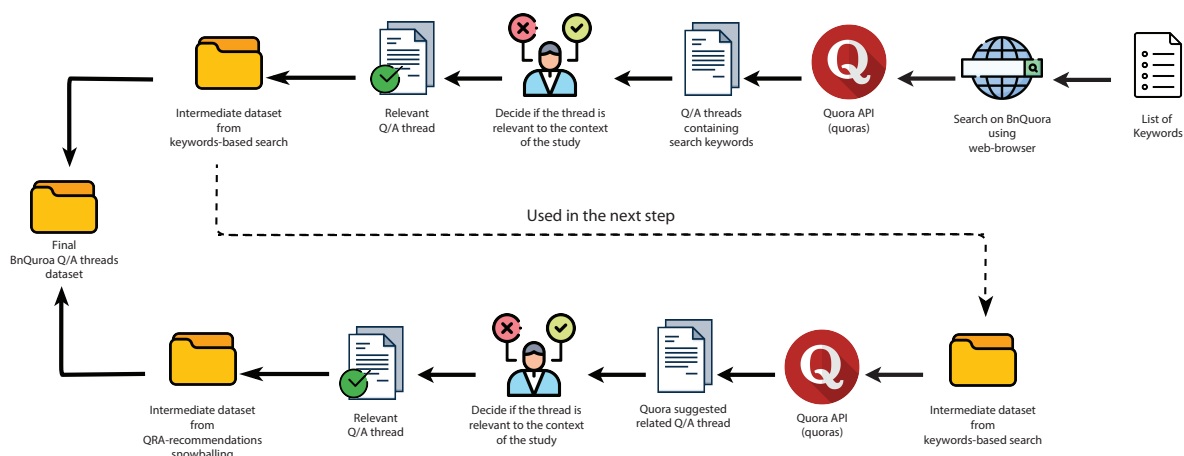


Figure 10.1: Trace data collection process in studies in BnQuora.

### 10.2.2.2 Strategies for Democratizing and Limiting Bias

I incorporated unpersonalized recommendations to broaden the scope, democratize, and limit researcher bias in data. While these recommendations can broaden the scope of datasets, their relevance should be considered before including them in the final datasets, where the data labeling process should be carefully thought out. A pictorial representation of the data collection process in both studies is shown in Figure 10.1.

**Algorithm as Apparatus: Broadening Scope through Unpersonalized Recommendation** I included the Quora recommendation algorithm (QRA)’s unpersonalized recommendations about related questions into their datasets as a form of snowball sampling [236]. As the QRA bases these recommendations on factors such as user co-visit data and popularity [614], they capture and reflect the aggregated interest and perceptions of the user mass on Q&A threads’ relevance to the topic of interest. Unlike the list of search keywords and phrases that reflected the sociohistoric understandings of the researchers or unlike the API that reflected the prior activities of an individual researcher, the QRA trains itself based on interaction data from a much larger number of users. Hence, the inclusion of QRA’s unpersonalized recommendations on related questions broadened the scope of the data. For example, in Chapter 5, while the term “Hindustan” (a colloquial endonym of “India”) was not present on their initial list of keywords, unpersonalized recommendations on

related questions retrieved several Q&A threads on this keyword [152]. Acting as an apparatus that reflects the opinions of Bengali Quora’s vast number of users, who themselves are members of colonially marginalized communities [149, 152]. Here, the apparatus and the phenomena co-configure each other. Thus, the documentary traces they generate are not pre-given but created, leading to diffractive approaches for understanding the data [423].

### **Data Labeling Workers as Apparatus: Balancing between Broader Scope and Bounded**

**Context** While unpersonalized diverse recommendations support exploring new ideas [108, 322], datasets need to be bounded within a context for research. I found that not all Q&A threads in the search results retrieved by using the list of keywords and phrases and the ones recommended by the QRA during their data collection process were relevant to the context of their study. For example, searching with the term “colonial” retrieved two particular questions—while one was related to British colonization, the other was about colonization on Mars. While the first question was relevant to the context of my study on Bengali communities’ identity work in relation to British colonization, the second was not considered relevant. Therefore, as an effort to contextualize the dataset, they used an additional data relevance labeling step between retrieving a potentially relevant Q&A thread (from keyword searching or QRA) and including that into the dataset.

Whereas in-depth qualitative or quantitative analysis of trace data requires critical thinking, expertise in specific methods (e.g., inductive thematic analysis, statistical models), tasks like labeling whether a data instance is related to the context of the study is less skill-intensive and are often done by crowdsourcing [305, 507]. However, the identities of communities or interpretations of individuals performing the data labeling tasks can shape the final dataset [488, 507]. As data labeling workers navigate through the fuzzy boundaries of phenomena of interest, their perspectives play a role in the inclusion/exclusion of trace data, and they collectively function as an apparatus through which researchers see and study the data.

### 10.2.3 Quantitative Methods in Critical Scholarship

The question of whether quantitative methods are inherently colonial is complex and multifaceted, touching on themes from the history of science, epistemology, and the sociology of knowledge [85]. Historically, the methods of research and those who have designed and used them have been complicit in the propagation of knowledge that has systematically promoted a pro-white and pro-Western agenda [444]. During the colonial era, scientific methods, including quantitative methods, were often used to justify and maintain colonial rule. For example, as explored by Moses and Knutsen [400], regression analysis, as the primary statistical method developed by and through Eugenics—the study of people’s physical traits—was complicit in the perpetuation of white superiority and Black inferiority across multiple physical dimensions. This becomes important when understood historically, as we have to consider that science, especially at that time, was largely driven by white cisgender individuals in the West. That very “science” has since contributed to how people are classified and categorized in their societies. These problematic classifications and categorizations have been deeply integrated into the social structures and systems that shape people’s everyday experiences, which have been especially harmful to People of Color [57].

Proponents of quantitative methods argue that these approaches are objective and neutral [180], designed to uncover truths about the world irrespective of the observer’s background. However, critics argue that these methods’ supposed objectivity can mask underlying biases, especially if the questions posed and colonial mindsets influence the interpretations of the data. Some researchers advocate for a critical approach to quantitative research, which involves being aware of and addressing the power dynamics and biases that can influence the research process and outcomes [325]. Scholars like Linda Tuhiwai Smith argue for the need to decolonize research methods, including quantitative approaches [528]. This involves critically examining the assumptions underpinning these methods and adapting them to be more inclusive of diverse epistemologies and worldviews. Therefore, while colonial histories and mindsets can influence quantitative methods’ development, application, and interpretation, these methods themselves are not inherently colonial.

Especially their critical application of these methods prioritizing local knowledge and needs can resist the colonial dynamics. It is crucial for researchers to critically engage with these methods, acknowledging and addressing their limitations and biases and striving toward more inclusive and equitable research practices. For example, in Chapter 9, I did not develop a quantitative methods-based system to moderate the colonially marginalized Bengali communities' interaction through sociotechnical systems. Instead, I used quantitative methods to examine the biases of algorithmic and data-centric systems, and I interpreted the results of those investigations drawing from post-colonial computing scholarship. Previously, I have also reflected on the universal expectations for qualitative research practices (e.g., sampling) and data collection approach in trace ethnography using various apparatuses [51, 67]. Next, I will reflect on the ethics of data curation.

#### 10.2.4 Ethical Consideration in Data Curation

In Chapter 9, I followed [55]'s guidelines for ethical considerations that recommend reflecting on curation rationales, language variety, demographic, and text characteristics, among other things. My identity and educational background put me in the capacity to privilege the agency of local communities in computing research, which is crucial in decolonizing language technology [70].

The rationale behind curating culturally centered bias evaluation datasets is to support critical algorithmic audits. BIBED facilitates so in Bengali computational linguistics research. Especially given its utility in studying fairness and bias and the language spoken by many native speakers of colonially marginalized and under-represented diverse identities, a Bengali identity-bias evaluation dataset is long overdue in the literature. I discussed my sociohistoric and cultural rationales behind focusing on gender, religion, and nationality earlier in the dissertation. However, building this dataset focusing on different identity dimensions within the under-represented Bengali community, the population can be subjected to a “visibility trap” [57] (e.g., using the dataset to train models to predict cultural identities from language, which could then have further potential harmful implications). On the one hand, this work brings people from the margins to the center and attempts to give voice to those who don't have it. On the other hand, simplifying complex human

identity across various dimensions for NLP algorithms to understand also risks reductionist representation, datafication, and surveillance. In that chapter, I have considered binary categories for different identity dimensions. By including female and male identities only, my presented dataset does not represent non-binary gender identity like হিজড়া (*/ˈɦidʒɽa/*, loosely corresponds to Western queer and transgender identities [415]) in Bengali communities. Again, though considering the Hindu and Muslim communities in the case of religion-based identity account for the large majority of the Bengali population, I recognize that religious minority Buddhist and Christian communities ( $\sim 1\%$ ) [92, 299] are excluded from the bias evaluation dataset. Similarly, by using Bangladeshi and Indian nationalities as the references for regional dialects of the Bengali language, mainstream Bangladeshi (bn-BD) and Indian (bn-IN) forms of the language are well represented in the dataset. However, I conflated and lost nuances for smaller regional dialects like Chittagorian [201] and excluded the Bengali diaspora of other nationalities. Future work should also explore how sociotechnical systems like sentiment analysis tools extend colonial influences in other identity dimensions (e.g., caste, sexuality) in Bengali communities.

It is important to highlight how, in many cases, it can be difficult to explore the nuances and fluidity of people’s gender and sexual expression as the tools and datasets often represent data in binary ways, or nuance can become lost when explored as aggregated data. While adhering to this binary notion of identity streamlined the experiment setup, this limitation of my work is indicative of the field’s limitations, in general—to be restricted to using artifacts produced in colonial ontologies as research materials.

In some stages of building the dataset, I sampled sentences from an existing dataset [262] collected from Wikipedia, encyclopedias, and classic literature. Here, the expectation is that the writers of those texts are native Bengali speakers. The list of common names and surnames of underprivileged caste Hindu communities was developed by Bengali researchers and governmental authorities [167, 600]. To address the concern of data colonialism [132, 562], I consciously avoided scrapping data from social media that users often do not anticipate to be used in research [209]. While using public test results for contemporary common male and female names in Hindu and

Muslim communities, I randomly combined first, middle, and last names from the list to protect people’s privacy. Due to the textual nature of the dataset, it does not address the regional variation in accent or pronunciation. Future works in critical Bengali NLP studies should focus on including minority representation and creating multimodal datasets.

Moreover, in this study, I examine BSA tools’ bias in relation to Bengali categorical identities within a single dimension, focusing on gender, religion, and nationality individually. Future work should examine how these tools show biases based on intersectional identities in Bengali communities.

### **10.3 Implications for HCI and CSCW Research**

In this section, I discuss the implications of this dissertation research by reflecting on the postcolonial sociomateriality of online interaction, urging for historicist sensibility in cross-cultural computing research, and calling for engineering activism in critical HCI.

#### **10.3.1 Postcolonial Sociomateriality of Online Interaction**

HCI research adopting a postcolonial lens—which argues that all design research is culturally located and power-laden [282]—often tends to overlook or exaggerate the agency of the users [329, 500]. To understand the agency of both human and non-human entities comprising the BnQuora platform, I forward the concept of “postcolonial sociomateriality.” The term “sociomaterial” means that information and text are not inherently virtual, but rather, the way the digital world is designed and constructed has material consequences on the physical world, influencing how people act and interact [71, 179]. Here, the sociomateriality of an object emerges when the agency of the users and the technology operate together in practice. Taken together, I define postcolonial sociomateriality as the dualism of how a community engages with a design being determined by their cultural epistemologies [282], and in turn, that engagement unfolding material consequences into the community [179].

Chapter 5 shows how the Quora platform is malleable in how people adapt it to support local practices across diverse intersecting social worlds, yet robust enough to maintain a common identity. As Quora expands to include many regional languages, its identity and functions as a Q&A site remain unchanged. However, it expands its features to other conditions, i.e., it becomes a linguistically accessible platform to different regional audiences. The question arises of how these different audiences can use these language-specific Q&A platforms. I described earlier that the users of BnQuora were building communities by translating local practices (e.g., addressing stranger users as *bhai/dada* or *apa/didi*) on the platform and creating dialogue around their sociohistoric backgrounds—a generative model of culture. They used the Q&A website to share their lived experiences, make sense of the marginalization of their identity by colonial influence, discuss their hopes and ideas about decolonization, and call for reformation in sociopolitical structures.

At this stage, I will reflect on the material subjectivities that are shaping these online discourses. How are these decolonial discourses connected to and played out beyond the online sphere? The creative discursive space that the YouTubers shape shapes their and their audiences' consciousness, as I saw in Chapters 7 and 8. It was evident through my participants' encounters in a foreign hotel, female YouTubers' fear of safety for content making, and most YouTubers receiving invitations through newly formed social relationships on YouTube based on linguistic similarities. The participants' journalistic efforts to identify local economic and industrial potential and report corruption resemble recent works that show how these movements that are initially organized on online platforms often bring material outcomes, such as by creating social relationships, increasing transparency, and causing policy changes within governments and organizations [259, 348]. Moreover, in the contemporary landscape of decolonization, scholars find technology to be more effective in executing decolonial actions and generative of real-world change [330].

While the sociohistoric background of the Bengali people influenced the underlying form, function, and uses of BnQuora, following the argument on ontological design [606]—everything that we design in turn designs us back—the discussion in this space too shapes the users' consciousness. In the action phase of decolonization, the platform users expressed their vision about reforming

the region's regional sociopolitical structures and economic policies. These discussions are not just descriptive; rather, they were also potentially generative of real-world change [184]. Besides shaping the identities of those who used the platform, I argue that the shared information and generated textual conversation on this virtual medium will eventually lead to material outcomes, i.e., the calls for reformation being executed in the physical world. Contrary to orientalism [473]—the lens through which the West viewed the East—which was a colonial construct, the resilient narratives that BnQuora users are generating on the platform is a decolonizing construct in the making. While I have seen BnQuora users talking about the already increasing cooperation between Bangladeshi and Indian media industries, a more impactful materialization of this decolonization work would be possible with policy changes at the governmental levels (e.g., making international travel easier and more accessible in the region). To understand whether and how such decolonization movements materialize beyond the scope of online platforms, I call for future studies to draw on a *postcolonial sociomateriality* lens.

### 10.3.2 Invitation to Critically Study Nationalism in CSCW

I develop a call to action for the critical study of nationalism in social computing research. Although nationalism plays a significant role in shaping people's collective identity, there has been a lack of critical examination of this topic within the field of social computing and HCI by identity scholars. Among social computing scholarship that has foregrounded the idea of nationalism, most have used nationalism as a typology in computational social science research to study topical focuses of political campaigns [429, 485] and hate speech (e.g., xenophobia, anti-Semitism, racism) [404, 420, 620] in social media. However, such use of nationalism as a classification in datasets does not speak to its complex and deep entanglements with social relationships and how those translate through different institutions and technologies. While some scholars have recently looked at how sociotechnical systems' designs reinforce platformed racism and how it relates to white nationalism and supremacy [250, 332, 610], to the best of my knowledge, my research delves into the complexity of nationalism to a deeper extent previously unexplored in CSCW and social computing scholarship.

I argue that there is a pressing need for further research on nationalism to deepen the understanding of its evolving manifestations, impacts, and interconnections with various socio-political dynamics. By exploring its complexities across diverse geopolitical and cultural contexts and time periods, we can examine the multifaceted nature of nationalism and its role in shaping identities and interactions in online communities, thus offering valuable insights for informed design, content moderation, and platform governance.

### 10.3.3 Historicist Understanding for Cross-Cultural Computing

In continuing to build on my study in Chapter 7, I ask the question of whether technology is nationalist and consider research as a kind of technology. Here, I situated historicism as an important avenue through which to continue investigating cross-cultural computing. A historical example of colonial influence on the understanding of culture is how colonial ontologies of culture categorized people in nation-states with artificial borders.

As computing as a field continues to bear colonial impulses, many cross-cultural computing scholars often adopt crude conceptualizations of cultures, a typical example of which is to use different nation-states, borders of which are colonially demarcated, as a proxy for different cultures [118, 141]. Even when they focus on the intersectional nuances introduced by age, gender, and socioeconomic status within the colonially marginalized communities [9, 297, 329, 392], their explorations often do not transcend the boundaries of one modern nation-state per study at a time or are structured as a comparative study of people's cultural perception in two or more nation-states (e.g., studying people's perception about hate speech or privacy in the US, China, Brazil, and India). There are also studies that focus on sub-cultures within the national border (e.g., language technologies' bias against African-American Vernacular English [476]). However, a dearth of research studied an ethnolinguistic or cultural group fissured and interconnected across the boundaries of nation-states—a pattern that disproportionately affected historical colonies and the Global South locations. Though colonial borders in Asia and Africa have historically separated many indigenous communities in different modern nation-states (e.g., Bengalis in Bangladesh and

India, Maasais in Kenya and Tanzania, and Punjabis in India and Pakistan), such transnational conceptualization of communities and how their technological practices were historically shaped are severely understudied in social computing.

While decolonial and postcolonial computing scholars focus on identifying colonial impulses of technology, often, they do not pay sufficient attention to the historical aspects of those societies and cultures. I argue that a historicist perspective is necessary to understand people's cultural expression and collective identity with nuances introduced by race, ethnicity, religion, and nationality. My study shows how YouTubers from three different countries, Bangladesh, India, and Pakistan, as well as two major religions, Hinduism and Islam, conceptualize collective local identity. I described how their individual and societal histories and sense of cultural affinity to particular imaginations of communities or Bengali national identity influenced them to make videos. For example, while Indian Bengalis perceive Bengali as an ethnolinguistic identity related to but separate from their country-based nationality, for Bangladeshi YouTubers, Bengali identity is intertwined with their Bangladeshi nationality. Such differences in how Indian and Bangladeshi Bengalis perceive their identity raise an intriguing question of whether Bengali identity should be viewed as a long-standing civilizational identity based on linguistic culture or a political one defined by nation-states. People's personal views on this question lead them to focus on different aspects of their lived experience, dimensions of cultural expression, and timelines for historical conversations. In addition to cultural expressions, I found that historical relationships among different countries often shape aspirations of decolonial narratives and political conversations on YouTube. Through my study, I respond to the recent invitation to adopt historicism in CSCW research [74, 534] and show how, instead of conceptualizing cultures using broad blanket proxies in computing, historicist perspectives can help us better understand colonially marginalized peoples, sociohistoric construction of their cultural identity, and their expression and activities online.

In this dissertation, I highlighted the multiple ways Bengali people conceptualize their imagined communities and nationhood in relation to the Indian subcontinent's colonial past. However, the importance of historicism is not unique to this geocultural context alone. Consider the case of

fraught race relationships in the US. Similar to how the British colonizers leaving the subcontinent did not end colonial division in Bengal but institutionally separated people using India-Pakistan borders, in the aftermath of the American Civil War, as the US entered the period of reconstruction, widespread racial segregation was institutionalized through Jim Crow laws [182]. A major impediment to socio-psychological decolonization in the Bengali context and annihilating racist ideologies in the American context is the hegemonic construction of convenient historical narratives. My participants talked about the prejudiced removal of historical figures from textbooks and the political and partisan control over Pakistani curriculum to downplay the oppression of the Bengali people and framing the Bangladeshi liberation war in 1971 as an Indian plot, and how they viewed these incomprehensive historical narratives obstructing the sociopolitical reformation of the region. Analogous to this, special interest groups like the Daughters of the Confederacy shaped the Southern textbooks through “lost cause” propaganda to make sure that the Southern soldiers were portrayed as heroes in schools and slavery was minimized as the primary cause of the Civil War [46]. To this day, these institutionally propagated narratives control much of public opinion and promote politicians who would promote racist logic, justify slavery, and uphold white nationalism.

#### **10.3.4 Call for Engineering Activism in Critical HCI**

By demonstrating the consistent discrepancies in fine-tuned BSA models’ and tools’ outputs for identical sentences expressing various identities, the work presented in Chapter 9 challenges the “functionality assumption” [447] of sentiment analysis that it relies on sentences’ lexical content and contexts and not on what identities they represent. In my discussion for that chapter, I explained the quantitative findings through a postcolonial understanding of the studied linguistic communities’ social, cultural, and historical contexts. This research, foregrounding the historically marginalized and under-represented Bengali community, contributes to the intersection of CHI, social computing, NLP, and fairness and bias literature contextualized in the Global South.

While critical HCI studies adopting a qualitative approach can provide deep and rich insights into biases in computational systems, those explorations are insufficient, and a fine-grained understanding of systems, architecture, algorithms, and code is essential for describing and explaining new information technologies' social, ethical, and political dimensions [413]. Building on that call for “engineering activism”—the use of engineering skills and knowledge to promote social justice, I argue that future NLP research (e.g., developing sentiment analysis tools), especially in critical HCI space, should actively reflect on identity-related biases and seek collaboration among individuals of diverse religious and transnational identities. Such socially aware engineering efforts can be particularly useful for an emerging effort in human-centered AI research that encourages “seamful designs” [190] of sociotechnical infrastructures for greater explainability and user agency. Instead of hiding seams—mismatches, gaps, cracks, and uncertainties that arise when a technology is deployed in the world or plainly treating them as problematic, strategically revealing and examining seams would support user agency, re-configuration, and appropriation. While recent works have often sought to trace and identify stages of AI dataset development [314, 490], through this work, I encourage the examination of “seams” (e.g., sentiment analysis components) of larger AI-based systems (e.g., automated content moderation). Moreover, I showed how a historicist sensibility [534], particularly understanding the sociocultural context in relation to coloniality [144], would better ground the engineering activist efforts.

## Chapter 11

### Conclusion

This dissertation presents studies from a broad and multi-platform investigation of the use of computing systems by colonially marginalized users in the Global South engaging in identity work. Focusing on the duality of sociotechnical systems both supporting and impeding the identity expression and performance of marginalized identities, I aimed to explore how people engage in the process of decolonizing their local and indigenous identities using computing technologies that perpetuate coloniality.

To pursue this objective, I focused on the colonially marginalized Bengali communities from South Asia. Though colonialism in the Bengal region of South Asia, in the sense of direct political control, ended in the last century, its legacy continues to mediate and shape the everyday lived experience of those whose local contexts were colonized. Online platforms, especially when designed to support regional languages (e.g., Bengali Quora), can become a space for colonized populations to decolonize their identities through conversations where they can more candidly and safely negotiate their intersectional identities across various dimensions (e.g., ethnic, linguistic, national, and religious). In this process, the idea of native or Indigenous identity is not monolithic but is complex and defined by pluralism. My research presents a nuanced account of how coloniality has shaped marginalized communities' ideas and experiences of nationalism. I described how individuals conceptualize their collective identities across various dimensions, along with their discernment of nationalism perpetuated by different institutions. I also highlight people's decolonial

strategies realized through collaborative textual discussions of everyday experiences in colonially shaped societies and video-mediated discourse in the forms of social interviews, decolonial storytelling, political explainers, and YouTube journalism. I discuss how these online interactions foster interconnected cultural understanding, construct comprehensive historical narratives, and foreground local and grassroots perspectives, which aim to strengthen and reaffirm native identities on social-psychological levels and facilitate material reformations in sociopolitical and economic spheres.

Though sociotechnical systems like online platforms can support identity performances and expression amongst marginalized populations, these spaces are also known to reinforce harmful ideologies and hegemonic values through various sociotechnical mechanisms. During this process of normalizing certain identities, the platform marginalizes other groups of users, manifesting a technological form of coloniality. In my research, I demonstrate how language makes identities subject to marginalization. My study highlights the role of social factors (e.g., postcolonial identities and relationships), human actors (e.g., moderation, majority user groups, developers), and technical components and policies (e.g., upvotes/downvotes, copyrights, and multi-channel network) in the governance, surveillance, monetization, and determination of visibility on various platforms (e.g., YouTube) in different scales. I also explored how the politics of algorithms and datasets continue to exacerbate the marginalization of identities in sociotechnical systems.

This interdisciplinary dissertation research seeks to understand the decolonization of identity and examine the coloniality of technology by employing a range of qualitative (e.g., trace ethnography, interviews) and quantitative (e.g., NLP, survey, experiment) methods and is informed by critical scholarship in decolonialism, postcolonialism, feminism, sociology, science and technology studies, political science, etc. In doing so, my research in HCI and CSCW contributes to the identity literature in the Global South, focuses on often-overlooked but crucial identity dimensions like religion and nationalism, investigates the genealogy of colonial marginalization of identities with a historicist sensibility, and explains users' strategies to negotiate their identities amid the challenges stemming from downstream colonial effects perpetuated by computing technologies and research

practices. Since colonization profoundly and prolongedly impacted people's identities that mediate their everyday interactions and experiences, it is absolutely important to understand people's practices around ICT platforms, explore the shortcomings of these technologies, generate design criteria, and develop systems to help support the decolonization of identity. There is yet a lot of work to be done and directions to be explored, and this dissertation contributes to this broad area of scholarship from the angles of human-computer interaction and social computing.

## Bibliography

- [1] Michael Adas. Imperialism and colonialism in comparative perspective. The International History Review, 20(2):371–388, 1998.
- [2] Rounak Afroz and Md Ataur Rahman. Transboundary river water for Ganges and Teesta rivers in Bangladesh: An assessment. Global Science and Technology Journal, 1(1):100–111, 2013.
- [3] The New Age. Hindu houses set on fire in Rangpur following facebook post. <https://www.newagebd.net/article/152086/hindu-houses-set-on-fire-in-rangpur-following-facebook-post>, 2021. Last accessed: Dec 4, 2022.
- [4] International Energy Agency. Emissions – global energy & CO2 status report 2019 – analysis - iea. <https://www.iea.org/reports/global-energy-co2-status-report-2019/emissions>, 2019. Last accessed: August 3, 2023.
- [5] Central Intelligence Agency (CIA) and Index Mundi. World languages - demographics. <https://www.indexmundi.com/world/languages.html>, 2020. Last accessed: July 2, 2021.
- [6] Samyak Agrawal, Kshitij Gupta, Devansh Gautam, and Radhika Mamidi. Towards detecting political bias in Hindi news articles. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop, pages 239–244, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [7] Sibbir Ahmad, Songqing Jin, Veronique Theriault, and Klaus Deininger. Labor market discrimination in Bangladesh: Experimental evidence from the job market of college graduates. Annual Meeting of Agricultural and Applied Economics Association, 2023.
- [8] Feroz Ahmed. The struggle in Bangladesh. Bulletin of Concerned Asian Scholars, 4(1):2–22, 1972.
- [9] Syed Ishtiaque Ahmed, Md Romael Haque, Jay Chen, and Nicola Dell. Digital privacy challenges with shared mobile phone use in Bangladesh. Proceedings of the ACM on Human-computer Interaction, 1(CSCW):1–20, 2017.
- [10] Morsheda Akhter and Philip Q Yang. The Bangladeshi diaspora in the united states: History and portrait. Genealogy, 7(4):81, 2023.

- [11] Olga Akselrod. How artificial intelligence can deepen racial and economic inequities. <https://www.aclu.org/news/privacy-technology/how-artificial-intelligence-can-deepen-racial-and-economicinequities>, 2021. Last accessed: Sep 11, 2023.
- [12] Uchenna Akujuobi and Xiangliang Zhang. Delve: a dataset-driven scholarly search and analysis system. ACM SIGKDD Explorations Newsletter, 19(2):36–46, 2017.
- [13] Ban Al-Ani, Gloria Mark, and Bryan Semaan. Blogging in a region of conflict: supporting transition to recovery. In Proceedings of the SIGCHI Conference on human factors in computing systems, pages 1069–1078, 2010.
- [14] Youssef Al Hariri, Walid Magdy, and Maria K Wolters. Atheists versus theists: religious polarisation in arab online communities. Proceedings of the ACM on human-computer interaction, 5(CSCW2):1–28, 2021.
- [15] M Alam. Poverty from the Wealth of Nations: Integration and Polarization in the Global Economy since 1760. Springer, 2016.
- [16] Mustafa Ali. Towards a decolonial computing. Ambiguous Technologies: Philosophical Issues, Practical Solutions, Human Nature, pages 28–35, 2014.
- [17] Syed Mustafa Ali. A brief introduction to decolonial computing. XRDS: Crossroads, The ACM Magazine for Students, 22(4):16–21, 2016.
- [18] Tariq Ali. Bangla desh: Results and prospects. New Left Review, 68:3–55, 1971.
- [19] Adriana Alvarado Garcia, Juan F Maestre, Manuhua Barcham, Marilyn Iriarte, Marisol Wong-Villacres, Oscar A Lemus, Palak Dudani, Pedro Reynolds-Cuéllar, Ruotong Wang, and Teresa Cerratto Pargman. Decolonial pathways: Our manifesto for a decolonizing agenda in hci research and design. In Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–9, 2021.
- [20] Bhimrao Ramji Ambedkar. National frustration. Dr. Babasaheb Ambedkar: Writings and Speeches, 1979.
- [21] Bhimrao Ramji Ambedkar. Annihilation of caste. Verso Books, 2014.
- [22] Bhimrao Ramji Ambedkar. Who were the Shudras. GENERAL PRESS, 2020.
- [23] Saleema Amershi, Maya Cakmak, William Bradley Knox, and Todd Kulesza. Power to the people: The role of humans in interactive machine learning. Ai Magazine, 35(4):105–120, 2014.
- [24] Tawfiq Ammari and Sarita Schoenebeck. Networked empowerment on facebook groups for parents of children with special needs. In Proceedings of the 33rd annual ACM conference on human factors in computing systems, pages 2805–2814, 2015.
- [25] Tawfiq Ammari and Sarita Schoenebeck. Understanding and supporting fathers and fatherhood on social media sites. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 1905–1914, 2015.

- [26] Tawfiq Ammari, Sarita Schoenebeck, and Meredith Morris. Accessing social support and overcoming judgment on social media among parents of children with special needs. In Proceedings of the international AAAI conference on web and social media, volume 8, pages 22–31, 2014.
- [27] Tawfiq Ammari, Sarita Schoenebeck, and Daniel M Romero. Pseudonymous parents: Comparing parenting roles and identities on the mommit and daddit subreddits. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pages 1–13, 2018.
- [28] Benedict Anderson. Imagined communities: Reflections on the origin and spread of nationalism. Verso books, 2006.
- [29] Robert J Anderson. Representations and requirements: The value of ethnography in system design. Human-computer interaction, 9(2):151–182, 1994.
- [30] Maddison Angus. Development centre studies the world economy historical statistics: historical statistics. OECD Publishing, 2003.
- [31] Ahmed Ansari. Design’s missing others and their incommensurate worlds. In Design in Crisis, pages 137–158. Routledge, 2020.
- [32] Cecilia Aragon, Shion Guha, Marina Kogan, Michael Muller, and Gina Neff. Human-centered data science: an introduction. MIT Press, 2022.
- [33] Ofer Arazy, Oded Nov, Raymond Patterson, and Lisa Yeo. Information quality in wikipedia: The effects of group composition and task conflict. Journal of Management Information Systems, 27(4):71–98, 2011.
- [34] Sabrina Argoub. The NLP divide: English is not the only natural language. <https://blogs.lse.ac.uk/polis/2021/06/09/the-nlp-divide-english-is-not-the-only-natural-language/>, 2021. Last accessed: Sep 10, 2023.
- [35] Gaurav Arora. inltk: Natural language toolkit for Indic languages. arXiv preprint arXiv:2009.12534, 2020.
- [36] Tina Askanus. Video for change. The handbook of development communication and social change, 2014.
- [37] Tina Askanus. Online video activism and political mash-up genres. Jomec Journal, 4, 2016.
- [38] Tina Askanus and Nils Gustafsson. Mainstreaming the alternative: The changing media practices of protest movements. Interface: a journal for and about social movements, 2009(2):23–41, 2010.
- [39] Tina Askanus and Julie Uldam. Online social media for radical politics: climate change activism on youtube. International journal of electronic governance, 4(1-2):69–84, 2011.
- [40] Imran Awan. Islamophobia on social media: A qualitative analysis of the facebook’s walls of hate. International Journal of Cyber Criminology, 10(1):1, 2016.
- [41] Kagonya Awori, Frank Vetere, and Wally Smith. Sessions with grandma: fostering indigenous knowledge through video mediated communication. In Proceedings of the First African Conference on Human Computer Interaction, pages 1–11, 2016.

- [42] Senthil Kumar B, Pranav Tiwari, Aman Chandra Kumar, and Aravindan Chandrabose. Casteism in India, but not racism - a study of bias in word embeddings of Indian languages. In Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference, pages 1–7, Marseille, France, June 2022. European Language Resources Association.
- [43] Younggwe Bae and Hongchul Lee. Sentiment analysis of twitter audiences: Measuring the positive or negative influence of popular twitterers. Journal of the American Society for Information Science and technology, 63(12):2521–2535, 2012.
- [44] Ricardo Baeza-Yates. Bias in search and recommender systems. In Proceedings of the 14th ACM Conference on Recommender Systems, pages 2–2, 2020.
- [45] Jasodhara Bagchi. Representing nationalism: Ideology of motherhood in colonial bengal. Economic and Political Weekly, pages WS65–WS71, 1990.
- [46] Fred Arthur Bailey. The textbooks of the” lost cause”: censorship and the creation of southern state histories. The Georgia Historical Quarterly, 75(3):507–533, 1991.
- [47] Sarbani Banerjee. “More or Less” Refugee?: Bengal Partition in Literature and Cinema. Univ. of Western Ontario, 2015.
- [48] Government of the People’s Republic of Bangladesh. The constitution of the people’s republic of Bangladesh: Nationalism, 1972. Last accessed: Aug 28, 2023.
- [49] Srijan Bansal, Vishal Garimella, Ayush Suhane, and Animesh Mukherjee. Debiasing multi-lingual word embeddings: A case study of three indian languages. In Proceedings of the 32nd ACM Conference on Hypertext and Social Media, pages 27–34, 2021.
- [50] Karen Barad. Meeting the universe halfway. In Meeting the universe halfway. Duke University Press, 2007.
- [51] Karen Barad. Diffracting diffraction: Cutting together-apart. Parallax, 20(3):168–187, 2014.
- [52] Jeffrey Bardzell and Shaowen Bardzell. “a great and troubling beauty”: cognitive speculation and ubiquitous computing. Personal and ubiquitous computing, 18(4):779–794, 2014.
- [53] Kristen Barta and Nazanin Andalibi. Constructing authenticity on tiktok: Social norms and social support on the” fun” platform. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–29, 2021.
- [54] Amita Baviskar. The politics of being ‘indigenous’. Indigeneity in India, pages 33–50, 2006.
- [55] Emily M Bender and Batya Friedman. Data statements for natural language processing: Toward mitigating system bias and enabling better science. Transactions of the Association for Computational Linguistics, 6:587–604, 2018.
- [56] Ruha Benjamin. Cultura obscura: race, power, and “culture talk” in the health sciences. American journal of law & medicine, 43(2-3):225–238, 2017.
- [57] Ruha Benjamin. Race After Technology: Abolitionist Tools for the New Jim Code. Polity Press, 2019.

- [58] Youcef Benkhedda. quora-scrapers. <https://pypi.org/project/quora-scrapers/>, 2020. PyPI: <https://pypi.org/project/quora-scrapers/>, last accessed: March 27, 2022.
- [59] Jeremy Bentham. Panopticon or the inspection house. T. Payne, 1791.
- [60] Bruce Berman. Control and crisis in colonial Kenya: The dialectic of domination. East African Publishers, 1992.
- [61] Marianne Bertrand and Sendhil Mullainathan. Are emily and greg more employable than lakisha and jamal? a field experiment on labor market discrimination. American economic review, 94(4):991–1013, 2004.
- [62] Homi K Bhabha. The location of culture. Routledge, 2012.
- [63] Gurminder K Bhambra. Postcolonial and decolonial dialogues. Postcolonial studies, 17(2):115–121, 2014.
- [64] Varuni Bhatia. Unforgetting Chaitanya: Vaishnavism and cultures of devotion in colonial Bengal. Oxford University Press, 2017.
- [65] Shaily Bhatt, Sunipa Dev, Partha Talukdar, Shachi Dave, and Vinodkumar Prabhakaran. Re-contextualizing fairness in NLP: The case of India. In Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers), pages 727–740, Online only, November 2022. Association for Computational Linguistics.
- [66] Abhik Bhattacharjee, Tahmid Hasan, Wasi Ahmad, Kazi Samin Mubasshir, Md Saiful Islam, Anindya Iqbal, M. Sohel Rahman, and Rifat Shahriyar. BanglaBERT: Language model pretraining and benchmarks for low-resource language understanding evaluation in Bangla. In Marine Carpuat, Marie-Catherine de Marneffe, and Ivan Vladimir Meza Ruiz, editors, Findings of the Association for Computational Linguistics: NAACL 2022, pages 1318–1327, Seattle, United States, July 2022. Association for Computational Linguistics.
- [67] Sayan Bhattacharyya, Peter Organisciak, and J Stephen Downie. A fragmentizing interface to a large corpus of digitized text:(post) humanism and non-consumptive reading via features. Interdisciplinary Science Reviews, 40(1):61–77, 2015.
- [68] Raj S Bhopal. Ethnicity, race, and health in multicultural societies: foundations for better epidemiology, public health, and health care. Oxford University Press, 2007.
- [69] Nicola J Bidwell. Decolonising in the gaps: Community networks and the identity of african innovation. In Re-imagining Communication in Africa and the Caribbean, pages 97–115. Springer, 2021.
- [70] Steven Bird. Decolonising speech and language technology. In Proceedings of the 28th international conference on computational linguistics, pages 3504–3519, 2020.
- [71] Pernille Bjørn and Carsten Østerlund. Sociomaterial-Design: Bounding technologies in practice. Springer, 2014.
- [72] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. Language (technology) is power: A critical survey of “bias” in NLP. arXiv preprint arXiv:2005.14050, 2020.

- [73] Su Lin Blodgett and Brendan O'Connor. Racial disparity in natural language processing: A case study of social media african-american english. arXiv preprint arXiv:1707.00061, 2017.
- [74] Susanne Bødker, Sarah Fox, Nicolas Lalone, Megh Marathe, and Robert Soden. (re) connecting history to the theory and praxis of hci, 2023.
- [75] Zacchaeus Bolaji. quorapy. <https://github.com/djunehor/quorapy>, 2019. PyPI: <https://pypi.org/project/quorapy/>, last accessed: June 18, 2020.
- [76] Carlo Emilio Bonferroni. The calculation of assurance from groups of tests. Studi in Onore del Professore Salvatore Ortu Carboni, 1935.
- [77] Anuja Bose. Hindutva and the politicization of religious identity in India. Journal of Peace, Conflict and Development, 13(1):5–35, 2009.
- [78] Sugata Bose and Ayesha Jalal. Modern South Asia: history, culture, political economy. Routledge, 2017.
- [79] Geoffrey C Bowker and Susan Leigh Star. Sorting things out: Classification and its consequences. MIT press, 2000.
- [80] Encyclopaedia Britannica. Bengali language. <https://www.britannica.com/topic/Bengali-language>, 1998. Year retrieved through the article history. Last accessed: December 27, 2020.
- [81] Encyclopaedia Britannica. Hindustani language. <https://www.britannica.com/topic/Hindustani-language>, 2022. Online, Last accessed: December 21, 2022.
- [82] Encyclopaedia Britannica. Indian subcontinent | map, countries, population and histories. <https://www.britannica.com/place/Indian-subcontinent>, 2022. Last accessed: March 21, 2023. Archive link: <https://web.archive.org/web/20230316134406/https://www.britannica.com/place/Indian-subcontinent>.
- [83] Encyclopedia Britannica. Citizen journalism | Benefits, Challenges & Impact — britannica.com. <https://www.britannica.com/topic/citizen-journalism>, 2018. [Accessed 18-Jul-2023].
- [84] Encyclopedia Britannica. Adivasi | Indigenous, Tribal, Marginalized — britannica.com. <https://www.britannica.com/topic/Adivasi>, 2021. [Accessed 14-Jul-2023].
- [85] Lucile H Brockway. Science and colonial expansion: the role of the british royal botanic gardens. American ethnologist, 6(3):449–465, 1979.
- [86] Meredith Broussard. Artificial unintelligence. MIT press, 2019.
- [87] Karida Brown and José Itzigsohn. The sociology of web du bois: Racialized modernity and the global color line, 2020.
- [88] Nina Brown, Thomas McIlwraith, and Laura Tubelle de González. Perspectives: An open introduction to cultural anthropology, volume 2300. American Anthropological Association, 2020.
- [89] Wendy Brown. Walled states, waning sovereignty. Princeton University Press, 2010.

- [90] Amy Bruckman. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the internet. Ethics and Information Technology, 4:217–231, 2002.
- [91] Amy S Bruckman. Should You Believe Wikipedia?: Online Communities and the Construction of Knowledge. Cambridge University Press, 2022.
- [92] Bangladesh Statistics Bureau BSB. Preliminary report on population and housing census 2022 : English version. [https://drive.google.com/file/d/1Vhn2t\\_PbEzo5-NDGBeoFJq4XCoSz0VKg/view](https://drive.google.com/file/d/1Vhn2t_PbEzo5-NDGBeoFJq4XCoSz0VKg/view), 2022. Last accessed: Feb 28, 2023.
- [93] Mary Bucholtz and Kira Hall. Identity and interaction: A sociocultural linguistic approach. Discourse studies, 7(4-5):585–614, 2005.
- [94] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Conference on fairness, accountability and transparency, pages 77–91. PMLR, 2018.
- [95] Jean Burgess and Joshua Green. YouTube: Online video and participatory culture. John Wiley & Sons, 2018.
- [96] Moira Burke and Robert Kraut. Using facebook after losing a job: Differential benefits of strong and weak ties. In Proceedings of the 2013 conference on Computer supported cooperative work, pages 1419–1430, 2013.
- [97] The Business Standard. Bangladesh-India border is among world’s most dangerous borders. <https://www.tbsnews.net/explainer/worlds-most-dangerous-borders-175099>, 2020. Last accessed: May 20, 2023.
- [98] Judith Butler. Excitable speech: A politics of the performative. Routledge, 1997.
- [99] Judith Butler. Performative agency. Journal of cultural economy, 3(2):147–161, 2010.
- [100] Judith Butler. Gender trouble: Feminism and the subversion of identity. routledge, 2011.
- [101] Hilary M Carey. God’s empire: religion and colonialism in the British world, c. 1801–1908. Cambridge Univ. Press, 2011.
- [102] Aimé Césaire. Discourse on colonialism. NYU Press, 2001.
- [103] Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon. I tube, you tube, everybody tubes: analyzing the world’s largest user generated content video system. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement, pages 1–14, 2007.
- [104] Bhumitra Chakma. Bangladesh-India relations: Sheikh hasina’s India-positive policy approach. 2012.
- [105] Anindita Chakrabarty. Migrant identity at the intersection of postcolonialism and modernity. Journal of Migration Affairs, 2(2):100–116, 2020.
- [106] Dipesh Chakrabarty. Remembered villages: representation of Hindu-Bengali memories in the aftermath of the partition. Economic and Political Weekly, pages 2143–2151, 1996.

- [107] Dipesh Chakrabarty. Provincializing Europe: postcolonial thought and historical difference. Princeton University Press, 2008.
- [108] Abhijnan Chakraborty, Saptarshi Ghosh, Niloy Ganguly, and Krishna P Gummadi. Optimizing the recency-relevance-diversity trade-offs in non-personalized news recommendations. Information Retrieval Journal, 22(5):447–475, 2019.
- [109] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. The internet’s hidden rules: An empirical study of Reddit norm violations at micro, meso, and macro scales. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–25, 2018.
- [110] Rachel Charlotte Smith, Heike Winschiers-Theophilus, Asnath Paula Kambunga, and Sarala Krishnamurthy. Decolonizing participatory design: Memory making in namibia. In Proceedings of the 16th Participatory Design Conference 2020-Participation (s) Otherwise-Volume 1, pages 96–106, 2020.
- [111] Partha Chatterjee. The nation and its fragments: Colonial and postcolonial histories. Princeton University Press, 1993.
- [112] Joya Chatterji. Bengal divided: Hindu communalism and partition, 1932-1947. Cambridge University Press, 2002.
- [113] Uttiya Chattopadhyay. The role of rivers on man and society reflected in literature: A critical study of a few selected river-oriented popular bengali novels in geographical perspective. In Fluvial Systems in the Anthropocene: Process, Response and Modelling, pages 77–86. Springer, 2022.
- [114] Le Chen, Ruijun Ma, Anikó Hannák, and Christo Wilson. Investigating the impact of gender on rank in resume search engines. In Proceedings of the 2018 chi conference on human factors in computing systems, pages 1–14, 2018.
- [115] Le Chen, Alan Mislove, and Christo Wilson. Peeking beneath the hood of uber. In Proceedings of the 2015 internet measurement conference, pages 495–508, 2015.
- [116] Si Chen, Xinyue Chen, Zhicong Lu, and Yun Huang. ” my culture, my people, my hometown”: Chinese ethnic minorities seeking cultural sustainability by video blogging. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1):1–30, 2023.
- [117] John Cheney-Lippold. We are data: Algorithms and the making of our digital selves. NYU Press, 2018.
- [118] Hichang Cho, Bart Knijnenburg, Alfred Kobsa, and Yao Li. Collective privacy management in social media: A cross-cultural validation. ACM Transactions on Computer-Human Interaction (TOCHI), 25(3):1–33, 2018.
- [119] Danielle Keats Citron. Hate crimes in cyberspace. Harvard University Press, 2014.
- [120] The City of Boulder and Xcel Energy. Energy partnership agreement: 2023 – 2025 overview. [bit.ly/BoulderEnergy2023](https://bit.ly/BoulderEnergy2023), 2023. Last accessed and archived: Jan 19, 2024.

- [121] Kevin Clark, Minh-Thang Luong, Quoc V Le, and Christopher D Manning. *Electra: Pre-training text encoders as discriminators rather than generators*. arXiv preprint arXiv:2003.10555, 2020.
- [122] Rachel Clarke, Peter Wright, Madeline Balaam, and John McCarthy. Digital portraits: photo-sharing after domestic violence. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 2517–2526, 2013.
- [123] Jacob Cohen. Statistical power analysis for the behavioral sciences. Academic press, 2013.
- [124] Jacob Cohen. *A power primer*. PubMed, 2016.
- [125] Stephen J Collier. Topologies of power: Foucault’s analysis of political government beyond ‘governmentality’. Theory, culture & society, 26(6):78–108, 2009.
- [126] Patricia Hill Collins. Black feminist thought: Knowledge, consciousness, and the politics of empowerment. routledge, 2022.
- [127] William J Conover and Ronald L Iman. On multiple-comparisons procedures. Technical Report, LA-7677-MS:124–129, 1979.
- [128] Dacca Consulate General. Telegram 978 from the consulate general in dacca to the department of state. <https://2001-2009.state.gov/documents/organization/48049.pdf>, 1971. Last accessed: June 27, 2023.
- [129] Wikipedia contributors. Hindi–urdu controversy. [https://en.wikipedia.org/wiki/Hindi-Urdu\\_controversy](https://en.wikipedia.org/wiki/Hindi-Urdu_controversy), 2022. Online; Last accessed December 21, 2022.
- [130] Charles Horton Cooley. Human nature and the social order. Transaction Publishers, 1992.
- [131] Copyright.gov. U.S. copyright office fair use index. <https://www.copyright.gov/fair-use/>, 2023. Last accessed: June 23, 2024.
- [132] Nick Couldry and Ulises A Mejias. Data colonialism: Rethinking big data’s relation to the contemporary subject. Television & New Media, 20(4):336–349, 2019.
- [133] Kate Crawford. The atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press, 2021.
- [134] YouTube Creators. Fair use - copyright on youtube. <https://www.youtube.com/watch?v=1PvjRIkwI18>, 2019. Last accessed: June 22, 2024.
- [135] YouTube Creators. Content id claims & dispute process: Manage & action claims in studio. <https://www.youtube.com/watch?v=ybmRMEJG6LY>, 2022. Last accessed: June 22, 2024.
- [136] Kimberlé Crenshaw. Demarginalizing the intersection of race and sex: A black feminist critique of antidiscrimination doctrine, feminist theory and antiracist politics. In Feminist legal theories, pages 23–51. Routledge, 2013.
- [137] Jonathon N Cummings, John B Lee, and Robert E Kraut. *Communication technology and friendship during the transition from high school to college.*, 2006.

- [138] Peter Cummings. Arguments for and against standardized mean differences (effect sizes). Archives of pediatrics & adolescent medicine, 165(7):592–596, 2011.
- [139] Paula Czarnowska, Yogarshi Vyas, and Kashif Shah. Quantifying social biases in nlp: A generalization and empirical comparison of extrinsic fairness metrics. Transactions of the Association for Computational Linguistics, 9:1249–1267, 2021.
- [140] Lincoln Dahlberg. The corporate colonization of online attention and the marginalization of critical communication? Journal of Communication Inquiry, 29(2):160–180, 2005.
- [141] Hua Dai and Prashant C Palvi. Mobile commerce adoption in china and the united states: a cross-cultural study. ACM SIGMIS Database: the DATABASE for Advances in Information Systems, 40(4):43–61, 2009.
- [142] The Daily Star: Tech & Startup Desk. How short video platforms are playing a role in preserving cultural heritage and revitalising traditions. <https://www.thedailystar.net/tech-startup/news/how-short-video-platforms-are-playing-role-preserving-cultural-heritage-and-revitalising-traditions-3353316>, June 22, 2023, 10:16 PM. Last accessed: June 26, 2023.
- [143] E Valentine Daniel, Henry Bernstein, and Tom Brass. Plantations, proletarians, and peasants in colonial Asia, volume 11. Psychology Press, 1992.
- [144] Dipto Das. Studying multi-dimensional marginalization of identity from decolonial and postcolonial perspectives. In Companion Publication of the 2023 Conference on Computer Supported Cooperative Work and Social Computing, pages 437–440, 2023.
- [145] Dipto Das and Anthony J Clark. Construct of sarcasm on social media platform. In 2019 IEEE international conference on humanized computing and communication (HCC), pages 106–113. IEEE, 2019.
- [146] Dipto Das, Dhvani Gandhi, and Bryan Semaan. Reimagining communities through transnational bengali decolonial discourse with youtube content creators. pages 1–35, 2024. Under review at ACM Conference on Computer-Supported Cooperative Work and Social Computing.
- [147] Dipto Das, Shion Guha, and Bryan Semaan. Toward cultural bias evaluation datasets: The case of bengali gender, religious, and national identity. In Proceedings of the First Workshop on Cross-Cultural Considerations in NLP (C3NLP), pages 68–83, 2023.
- [148] Dipto Das, A.K.M. Najmul Islam, S.M. Taiabul Haque, Jukka Vuorinen, and Syed Ishtiaque Ahmed. Understanding the strategies and practices of facebook microcelebrities for engaging in sociopolitical discourses. In International Conference on Information & Communication Technologies and Development, pages 1–19, 2022.
- [149] Dipto Das, Carsten Østerlund, and Bryan Semaan. “jol” or “pani”? : How does governance shape a platform’s identity? Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–25, 2021.
- [150] Dipto Das, Arpon Podder, and Bryan Semaan. Note: A sociomaterial perspective on trace data collection: Strategies for democratizing and limiting bias. In ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS), pages 569–573, 2022.

- [151] Dipto Das and Bryan Semaan. quoras: A python api for quora data collection to increase multi-language social science research. In Conference Companion Publication of the 2020 on Computer Supported Cooperative Work and Social Computing, pages 251–256, 2020.
- [152] Dipto Das and Bryan Semaan. Collaborative identity decolonization as reclaiming narrative agency: Identity work of bengali communities on quora. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pages 1–23, 2022.
- [153] Dipto Das and Bryan Semaan. Decolonial and postcolonial computing research: A scientometric exploration. In Companion Publication of the 2022 Conference on Computer Supported Cooperative Work and Social Computing, pages 168–174, 2022.
- [154] Mithun Das and Animesh Mukherjee. Banglaabusememe: A dataset for bengali abusive meme classification. arXiv preprint arXiv:2310.11748, 2023.
- [155] Mitra Das. Internal colonialism and the movement for Bangladesh: A sociological analysis. In The Neon Electric Saraswati, pages 93–104. Brill, 1978.
- [156] Veena Das. Life and Words: Violence and the Descent into the Ordinary. Univ of California Press, 2006.
- [157] Saloni Dash, Rynaa Grover, Gazal Shekhawat, Sukhnidh Kaur, Dibyendu Mishra, and Joyojeet Pal. Insights into incitement: A computational perspective on dangerous speech on twitter in India. In ACM SIGCAS/SIGCHI Conference on Computing and Sustainable Societies (COMPASS), pages 103–121, 2022.
- [158] Débora de Castro Leal, Ana Maria Bustamante Duarte, Max Krüger, and Angelika Strohmayr. Into the mine: Wicked reflections on decolonial thinking and technologies. In C&T’21: Proceedings of the 10th International Conference on Communities & Technologies-Wicked Problems in the Age of Tech, pages 269–280, 2021.
- [159] Nathalie De Fabrique, Stephen J Romano, Gregory M Vecchi, and Vincent B Van Hasselt. Understanding stockholm syndrome. FBI L. Enforcement Bull., 76:10, 2007.
- [160] Jeroen JH Dekker. The fragile relation between normality and marginality. marginalization and institutionalization in the history of education. Peadagogica Historica, 26(2):12–29, 1990.
- [161] Maria del Pilar Kaladeen. ” those not with us anymore”: The literary archive of Indian minorities in Guyanese indenture and beyond. Journal of West Indian Literature, 29(1):26–191, 2021.
- [162] Michael A DeVito, Ashley Marie Walker, and Jeremy Birnholtz. ’too gay for facebook’ presenting lgbtq+ identity throughout the personal social media ecosystem. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–23, 2018.
- [163] Faisal Devji. Muslim Zion: Pakistan as a political idea. Hurst Publishers, 2013.
- [164] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805, 2018.

- [165] Amrit Dhillon. “no dogs or Indians”: Colonial britain still rules at India’s private clubs. <https://www.smh.com.au/world/no-dogs-or-indians-colonial-britain-still-rules-at-indias-private-clubs-20170630-gx1vtk.html>, 2017.
- [166] Mark Díaz, Isaac Johnson, Amanda Lazar, Anne Marie Piper, and Darren Gergle. Addressing age-related bias in sentiment analysis. In Proceedings of the 2018 chi conference on human factors in computing systems, pages 1–14, 2018.
- [167] Afia Dil. The Hindu and Muslim Dialects of Bengali. Stanford University, 1972.
- [168] Paul Dimeo. Colonial bodies, colonial sport: ’martial’punjabis, ’effeminate’bengalis and the development of Indian football. The international journal of the history of sport, 19(1):72–90, 2002.
- [169] Xianghua Ding, Yubo Kou, Yiwen Xu, and Peng Zhang. “as uploaders, we have the responsibility”: Individualized professionalization of bilibili uploaders. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2022.
- [170] divinAI. Diversity in artificial intelligence: Acm facct 2020. <https://divinai.org/conf/74/acm-facct>, 2020. Last accessed: Sep 12, 2023.
- [171] Lynn Dombrowski, Ellie Harmon, and Sarah Fox. Social justice-oriented interaction design: Outlining key design strategies and commitments. In Proceedings of the 2016 ACM Conference on Designing Interactive Systems, pages 656–671, 2016.
- [172] Bryan Dosono. Identity Work of Asian Americans and Pacific Islanders on Reddit: Traversals of Deliberation, Moderation, and Decolonization. PhD thesis, Syracuse University, 2019.
- [173] Bryan Dosono and Bryan Semaan. Identity work as deliberation: Aapi political discourse in the 2016 us presidential election. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pages 1–12, 2018.
- [174] Bryan Dosono and Bryan Semaan. Moderation practices as emotional labor in sustaining online communities: The case of aapi identity work on Reddit. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–13, 2019.
- [175] Bryan Dosono and Bryan Semaan. Decolonizing tactics as collective resilience: Identity work of aapi communities on Reddit. Proceedings of the ACM on Human-Computer interaction, 4(CSCW1):1–20, 2020.
- [176] Evelyn Douek. Facebook’s oversight board: Move fast with stable infrastructure and humility. NCJL & Tech., 21:1, 2019.
- [177] Paul Dourish. Technical and social features of categorization schemes. In Schmidt, K.; Simone, C.; Star, SL: Workshop Classification Schemes. Citeseer, 2000.
- [178] Paul Dourish. Implications for design. In Proceedings of the SIGCHI conference on Human Factors in computing systems, pages 541–550, 2006.
- [179] Paul Dourish. The stuff of bits: An essay on the materialities of information. MIT Press, 2017.

- [180] Paul Dourish and Scott D Mainwaring. UbiComp’s colonial impulse. In Proceedings of the 2012 ACM conference on ubiquitous computing, pages 133–142, 2012.
- [181] William Edward Burghardt Du Bois. The Souls of Black Folk: Essays and Sketches. A. C. McClurg, 1909.
- [182] William Edward Burghardt Du Bois. Black Reconstruction in America: Toward a History of the Part Which Black Folk Played in the Attempt to Reconstruct Democracy in America, 1860-1880. Reprint, Transaction Publishers, 1935.
- [183] Edward Duca. Malta stockholm syndrome (or why we love the british). 2015.
- [184] Robert Dunham. The generative foundations of action in organizations: Speaking and listening. International Journal of Coaching in Organizations, 7(2):43–63, 2009.
- [185] Sutapa Dutta. Packing a punch at the bengali babu. South Asia: Journal of South Asian Studies, 44(3):437–458, 2021.
- [186] Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In Proceedings of the 3rd innovations in theoretical computer science conference, pages 214–226, 2012.
- [187] Brianna Dym, Jed R Brubaker, Casey Fiesler, and Bryan Semaan. ” coming out okay” community narratives for lgbtq identity recovery work. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–28, 2019.
- [188] Benjamin Edelman, Michael Luca, and Dan Svirsky. Racial discrimination in the sharing economy: Evidence from a field experiment. American economic journal: applied economics, 9(2):1–22, 2017.
- [189] Benjamin G Edelman and Michael Luca. Digital discrimination: The case of airbnb. com. Harvard Business School NOM Unit Working Paper, 14-054, 2014.
- [190] Upol Ehsan, Q Vera Liao, Samir Passi, Mark O Riedl, and Hal Daume III. Seamful xai: Operationalizing seamful design in explainable ai. arXiv preprint arXiv:2211.06753, 2022.
- [191] Upol Ehsan and Mark O Riedl. Human-centered explainable ai: Towards a reflective sociotechnical approach. In HCI International 2020-Late Breaking Papers: Multimodality and Intelligence: 22nd HCI International Conference, HCII 2020, Copenhagen, Denmark, July 19–24, 2020, Proceedings 22, pages 449–466. Springer, 2020.
- [192] Sheena Erete, Aarti Israni, and Tawanna Dillahunt. An intersectional approach to designing in the margins. Interactions, 25(3):66–69, 2018.
- [193] Erik H Erikson. Identity: Youth and crisis, volume 7. WW Norton & company, 1968.
- [194] Maria Eriksson and Anna Johansson. Tracking gendered streams. Culture unbound. Journal of Current Cultural Research, 9(2):163–183, 2017.
- [195] Charlotte Eubanks and Pasang Yangjee Sherpa. We are (are we?) all indigenous here, and other claims about space, place, and belonging in asia. Verge: Studies in Global Asias, 4(2):vi–xiv, 2018.

- [196] Oliver Falck, Stephan Heblich, Alfred Lameli, and Jens Südekum. Dialects, cultural identity, and economic exchange. *Journal of urban economics*, 72(2-3):225–239, 2012.
- [197] Angela Fan and Claire Gardent. Generating biographies on Wikipedia: The impact of gender bias on the retrieval-based generation of women biographies. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 8561–8576, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [198] Frantz Fanon. *The Wretched of the Earth*. Grove Atlantic, 2007.
- [199] Frantz Fanon. *Black skin, white masks*. Grove press, 2008.
- [200] A.B.M. Razaul Karim Faquire. Language situation in Bangladesh. *The Dhaka University Studies*, pages 63–77, 2010.
- [201] A.B.M. Razaul Karim Faquire. On the classification of varieties of bangla spoken in Bangladesh. *Bup Journal*, 1(1):130–139, 2012.
- [202] Marta Fernández and Paulo Esteves. Silencing colonialism: Foucault and the international. In *Foucault and the Modern International*, pages 137–153. Springer, 2017.
- [203] Casey Fiesler. Lawful users: Copyright circumvention and legal constraints on technology use. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–11, 2020.
- [204] Casey Fiesler and Amy Bruckman. Copyright terms in online creative communities. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2551–2556. ACM New York, NY, USA, 2014.
- [205] Casey Fiesler and Amy S Bruckman. Creativity, copyright, and close-knit communities: A case study of social norm formation and enforcement. *Proceedings of the ACM on Human-Computer Interaction*, 3(GROUP):1–24, 2019.
- [206] Casey Fiesler, Jessica L Feuston, and Amy S Bruckman. Understanding copyright law in online creative communities. In *Proceedings of the 18th ACM conference on computer supported cooperative work & social computing*, pages 116–129, 2015.
- [207] Casey Fiesler, Jialun Jiang, Joshua McCann, Kyle Frye, and Jed Brubaker. Reddit rules! characterizing an ecosystem of governance. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12, 2018.
- [208] Casey Fiesler, Joshua Paup, and Corian Zacher. Chilling tales: Understanding the impact of copyright takedowns on transformative content creators. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–21, 2023.
- [209] Casey Fiesler and Nicholas Proferes. “participant” perceptions of twitter research ethics. *Social Media+ Society*, 4(1):2056305118763366, 2018.
- [210] Claudia Flores-Saviaga, Brian Keegan, and Saiph Savage. Mobilizing the trump train: Understanding collective action in a political trolling community. In *Proceedings of the International AAAI Conference on Web and Social Media*, pages 82–91, 2018.

- [211] Simon Fong, Yan Zhuang, Jinyan Li, and Richard Khoury. Sentiment analysis of online news using *mallet*. In 2013 International Symposium on Computational and Business Intelligence, pages 301–304. IEEE, 2013.
- [212] Michel Foucault. Discipline and punish: The birth of the prison. Vintage, 2012.
- [213] Mary Anne Franks. Revenge porn reform: A view from the front lines. Fla. L. Rev., 69:1251, 2017.
- [214] Nancy Fraser. Rethinking the public sphere: A contribution to the critique of actually existing democracy. Social text, 25/26:56–80, 1990.
- [215] Batya Friedman and Helen Nissenbaum. Bias in computer systems. ACM Transactions on Information Systems (TOIS), 14(3):330–347, 1996.
- [216] Roland G Fryer Jr and Steven D Levitt. The causes and consequences of distinctively black names. The Quarterly Journal of Economics, 119(3):767–805, 2004.
- [217] CJ Fuller. Ethnographic inquiry in colonial India: Herbert risley, william crooke, and the study of tribes and castes. Journal of the Royal Anthropological Institute, 23(3):603–621, 2017.
- [218] Sumit Ganguly. Explaining the Kashmir insurgency: political mobilization and institutional decay. International Security, 21(2):76–107, 1996.
- [219] Šumit Ganguly. Conflict unending: India-Pakistan tensions since 1947. Columbia University Press, 2002.
- [220] Diether Gebert, Sabine Boerner, and Debrabata Chatterjee. Do religious differences matter? an analysis in India. Team Performance Management: An International Journal, 2011.
- [221] Viktor Gecas. The self-concept. Annual review of sociology, 8(1):1–33, 1982.
- [222] R Stuart Geiger and David Ribes. Trace ethnography: Following coordination through documentary practices. In 2011 44th Hawaii international conference on system sciences, pages 1–10. IEEE, 2011.
- [223] Cherian George. Managing the dangers of online hate speech in south asia. Media Asia, 42(3-4):144–156, 2015.
- [224] Ysabel Gerrard. Beyond the hashtag: Circumventing content moderation on social media. New Media & Society, 20(12):4492–4511, 2018.
- [225] Anindita Ghoshal. ‘mirroring the other’: Refugee, homeland, identity and diaspora. In Routledge Handbook of Asian Diaspora and Development, pages 147–158. Routledge, 2021.
- [226] Sucheta Ghoshal and Amy Bruckman. The role of social computing technologies in grassroots movement building. ACM Transactions on Computer-Human Interaction (TOCHI), 26(3):1–36, 2019.
- [227] Sucheta Ghoshal, Rishma Mendhekar, and Amy Bruckman. Toward a grassroots culture of technology practice. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1):1–28, 2020.

- [228] Eric Gilbert. Widespread underprovision on Reddit. In Proceedings of the 2013 conference on Computer supported cooperative work, pages 803–808, 2013.
- [229] Sarah A Gilbert. ” i run the world’s largest historical outreach project and it’s on a cesspool of a website.” moderating a public scholarship site on Reddit: A case study of r/askhistorians. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1):1–27, 2020.
- [230] Michael Gillan. Refugees or infiltrators? the bharatiya janata party and “illegal” migration from Bangladesh. Asian studies review, 26(1):73–95, 2002.
- [231] Tarleton Gillespie. The politics of ‘platforms’. New media & society, 12(3):347–364, 2010.
- [232] Tarleton Gillespie. Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media. Yale University Press, 2018.
- [233] Barney G Glaser and Anselm L Strauss. Discovery of grounded theory: Strategies for qualitative research. Routledge, 2017.
- [234] Erving Goffman. The presentation of self in everyday life. Harmondsworth London, 1978.
- [235] Sunir Gohil, Sabine Vuik, Ara Darzi, et al. Sentiment analysis of health care tweets: review of the methods used. JMIR public health and surveillance, 4(2):e5789, 2018.
- [236] Leo A Goodman. Snowball sampling. The annals of mathematical statistics, pages 148–170, 1961.
- [237] Legislative Government of Bangladesh and Parliamentary Affairs Division. The constitution of the people’s republic of Bangladesh. <http://bdlaws.minlaw.gov.bd/act-367.html>, 1972. Last accessed: December 27, 2020.
- [238] Commissioner for Linguistic Minorities Government of India. Report of the commissioner for linguistic minorities: 50th report (july 2012 to june 2013). <https://web.archive.org/web/20160708012438/http://nclm.nic.in/shared/linkimages/NCLM50thReport.pdf>, 2014. Last accessed: December 27, 2020.
- [239] Antonio Gramsci. Selections from the prison notebooks of Antonio Gramsci. Elec Book, 1999.
- [240] Antonio Gramsci. The intellectuals. Contemporary sociological thought. Themes and theories, pages 49–58, 2005.
- [241] Sherri Grasmuck, Jason Martin, and Shanyang Zhao. Ethno-racial identity displays on facebook. Journal of computer-mediated communication, 15(1):158–188, 2009.
- [242] Franklin Graves and Michael Lee. The law of youtubers: The next generation of creators and the legal issues they face. Landslide, 9:8, 2016.
- [243] Jonathan D Greenberg. Generations of memory: remembering partition in India/pakistan and israel/palestine. Comparative Studies of South Asia, Africa and the Middle East, 25(1):89–110, 2005.
- [244] The Guardian. Seven dead after violence erupts during Hindu festival in Bangladesh. <https://www.theguardian.com/world/2021/oct/16/four-die-after-violence-erupts-at-hindu-festival-in-bangladesh>, 2021. Last accessed: Nov 4, 2022.

- [245] Ramachandra Guha. India after Gandhi: The history of the world's largest democracy. Pan Macmillan, 2017.
- [246] Ranajit Guha. Subaltern Studies VI: Writings on South Asian History. Oxford University Press, 1989.
- [247] Ranajit Guha. Dominance without hegemony: History and power in colonial India. Harvard University Press, 1997.
- [248] Ranajit Guha and Gayatri Chakravorty Spivak. Selected subaltern studies. Oxford University Press, 1988.
- [249] Jürgen Habermas. Moral consciousness and communicative action. MIT press, 1990.
- [250] Loni Hagen, Mary Falling, Oleksandr Lisnichenko, AbdelRahim A Elmadany, Pankti Mehta, Muhammad Abdul-Mageed, Justin Costakis, and Thomas E Keller. Emoji use in twitter white nationalism communication. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, 2019.
- [251] Zaglul Haider. A revisit to the Indian role in the Bangladesh liberation war. Journal of Asian and African Studies, 44(5):537–551, 2009.
- [252] Oliver L Haimson, Jed R Brubaker, Lynn Dombrowski, and Gillian R Hayes. Disclosure, stress, and support during gender transition on facebook. In Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing, pages 1176–1190, 2015.
- [253] Oliver L Haimson, Bryan Semaan, Brianna Dym, Joey Chiao-Yin Hsiao, Daniel Herron, and Wendy Moncur. Life transitions and social technologies: Research and design for times of life change. In Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, pages 480–486, 2019.
- [254] Helen Halbert and Lisa P Nathan. Designing for discomfort: Supporting critical reflection through interactive tools. In Proceedings of the 18th ACM conference on computer supported cooperative work & social computing, pages 349–360, 2015.
- [255] Stuart Hall. Cultural identity and cinematic representation. Journal of Cinema and Media, 1989.
- [256] Alex Hanna, Emily Denton, Andrew Smart, and Jamila Smith-Loud. Towards a critical race methodology in algorithmic fairness. In Proceedings of the 2020 conference on fairness, accountability, and transparency, pages 501–512, 2020.
- [257] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. Measuring price discrimination and steering on e-commerce web sites. In Proceedings of the 2014 conference on internet measurement conference, pages 305–318, 2014.
- [258] Anikó Hannák, Claudia Wagner, David Garcia, Alan Mislove, Markus Strohmaier, and Christo Wilson. Bias in online freelance marketplaces: Evidence from taskrabbit and fiverr. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, pages 1914–1933, 2017.

- [259] Karin Hansson, Malin Sveningsson, and Hillevi Ganetz. Organizing safe spaces:# metoo activism in sweden. *Computer Supported Cooperative Work (CSCW)*, 30(5):651–682, 2021.
- [260] Donna J Haraway and Thyrza Nichols Goodeve. *Modest\_witness@ second\_millennium*. routledge, 2018.
- [261] Drew Harwell. Why some accents don’t work on alexa or google home - washington post. <https://www.washingtonpost.com/graphics/2018/business/alexa-does-not-understand-your-accent/>, 2018. Last accessed: Sep 4, 2023.
- [262] Tahmid Hasan, Abhik Bhattacharjee, Kazi Samin, Masum Hasan, Madhusudan Basak, M Sohel Rahman, and Rifat Shahriyar. Not low-resource anymore: Aligner ensembling, batch filtering, and new datasets for bengali-english machine translation. *arXiv preprint arXiv:2009.09359*, 2020.
- [263] Daniel Hershcovich, Stella Frank, Heather Lent, Miryam de Lhoneux, Mostafa Abdou, Stephanie Brandl, Emanuele Bugliarello, Laura Cabello Piqueras, Ilias Chalkidis, Ruixiang Cui, et al. Challenges and strategies in cross-cultural nlp. *arXiv preprint arXiv:2203.10020*, 2022.
- [264] Danula Hettiachchi and Jorge Goncalves. Towards effective crowd-powered on-line content moderation. In *Proceedings of the 31st Australian Conference on Human-Computer-Interaction*, pages 342–346, 2019.
- [265] Michael A Hogg and Deborah I Terry. Social identity and self-categorization processes in organizational contexts. *Academy of management review*, 25(1):121–140, 2000.
- [266] Awal Hossain. Future of Bangladesh-India relationship-a critical analysis. Technical report, Air War College Montgomery United States, 2016.
- [267] Ishtiaq Hossain. Bangladesh-India relations: the ganges water-sharing treaty and beyond. *Asian Affairs: An American Review*, 25(3):131–150, 1998.
- [268] Sabine Hotho. Professional identity–product of structure, product of choice. *Journal of Organizational Change Management*, 2008.
- [269] Lara Houston, Steven J Jackson, Daniela K Rosner, Syed Ishtiaque Ahmed, Meg Young, and Laewoo Kang. Values in repair. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 1403–1414, 2016.
- [270] Dirk Hovy, Federico Bianchi, and Tommaso Fornaciari. “you sound just like your father” commercial machine translation systems include stylistic biases. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 1686–1690, 2020.
- [271] Tenghao Huang, Faeze Brahman, Vered Shwartz, and Snigdha Chaturvedi. Uncovering implicit gender bias in narratives through commonsense inference. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3866–3873, Punta Cana, Dominican Republic, November 2021. Association for Computational Linguistics.
- [272] Matthew Hughey. *The white savior film: Content, critics, and consumption*. Temple University Press, 2014.

- [273] Pere-Lluís Huguet Cabot, David Abadi, Agneta Fischer, and Ekaterina Shutova. Us vs. them: A dataset of populist attitudes, news bias and emotions. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, pages 1921–1945, Online, April 2021. Association for Computational Linguistics.
- [274] Margaret Hunter. The persistent problem of colorism: Skin tone, status, and inequality. Sociology compass, 1(1):237–254, 2007.
- [275] Herminia Ibarra and Jennifer L Petriglieri. Identity work and play. Journal of Organizational Change Management, 23(1):10–25, 2010.
- [276] U.S. Embassy in Bangladesh. More than 13,000 Bangladeshis—an all-time high—are currently studying in the united states during a record-setting year. <https://bd.usembassy.gov/more-than-13000-bangladeshis-an-all-time-high-are-currently-studying-in-the-united-states-during-a-record-setting-year/>, 2023. Last accessed: May 31, 2024.
- [277] Quora Inc. Why quora exists. <https://www.quora.com/about>, 2009. Last accessed: December 23, 2020.
- [278] DNA India. Assam government withdraws assamese as official language in barak valley, restores bengali. <https://www.dnaindia.com/india/report-assam-government-withdraws-assamese-as-official-language-in-barak-valley-restores-bengali-2017504>, 2014. Last accessed: December 27, 2020.
- [279] Times of India. International literacy day 2020: How literacy rate is calculated in India. <https://timesofindia.indiatimes.com/india/international-literacy-day-2020-how-literacy-rate-is-calculated-in-india/articleshow/77991200.cms>, 2020.
- [280] Global Media Insight. YouTube Statistics 2023 [Users by Country + Demographics] — globalmediainsight.com. <https://www.globalmediainsight.com/blog/youtube-users-statistics/>, 2023. [Accessed 12-Jul-2023].
- [281] Lilly Irani. Chasing innovation: Making entrepreneurial citizens in modern India. Princeton University Press, 2019.
- [282] Lilly Irani, Janet Vertesi, Paul Dourish, Kavita Philip, and Rebecca E Grinter. Postcolonial computing: a lens on design and development. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 1311–1320, 2010.
- [283] Lilly C Irani and Paul Dourish. Postcolonial interculturality. In Proceedings of the 2009 international workshop on Intercultural collaboration, pages 249–252, 2009.
- [284] Judy M Iseke-Barnes. Pedagogies for decolonizing. Canadian Journal of Native Education, 31(1):123–148, 2008.
- [285] Alvi Md Ishmam and Sadia Sharmin. Hateful speech detection in public Facebook pages for the Bengali language. In 2019 18th IEEE international conference on machine learning and applications (ICMLA), pages 555–560. IEEE, 2019.

- [286] Saman Ismail, Samiya Loya, and Jawed Hussain. Obsession for fair skin color in pakistan. International Journal of Innovation and Research in Educational Sciences, 2(6):2349–5219, 2015.
- [287] Steven J Jackson, Syed Ishtiaque Ahmed, and Md Rashidujjaman Rifat. Learning, innovation, and sustainability among mobile phone repairers in dhaka, Bangladesh. In Proc. of the 2014 conference on Designing interactive systems, 2014.
- [288] Christophe Jaffrelot. Minorities under attack in pakistan. In Minorities and Populism—Critical Perspectives from South Asia and Europe, pages 173–182. Springer, 2020.
- [289] Ayesha Jalal. Democracy and authoritarianism in South Asia. Cambridge University Press, 2009.
- [290] William James. The principles of psychology, volume 1. Cosimo, Inc., 2007.
- [291] Edmund Jan, Rupert Lee, Edmund Jan Osmańczyk, et al. Encyclopedia of the United Nations and International Agreements. Taylor & Francis, 2003.
- [292] Sadik Al Jarif. Bangla and english fake news detection dataset. <https://www.kaggle.com/datasets/sadikaljarif/bangla-and-english-fake-news-dataset>, 2022. Last accessed: Dec 25, 2023.
- [293] Patricia Jeffery and Amrita Basu. Appropriating gender: Women’s activism and politicized religion in South Asia. Routledge, 2012.
- [294] Shagun Jhaver, Iris Birman, Eric Gilbert, and Amy Bruckman. Human-machine collaboration for content regulation: The case of reddit automoderator. ACM Transactions on Computer-Human Interaction (TOCHI), 26(5):1–35, 2019.
- [295] Shagun Jhaver, Amy Bruckman, and Eric Gilbert. Does transparency in moderation really matter? user behavior after content removal explanations on reddit. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–27, 2019.
- [296] Jialun Aaron Jiang, Peipei Nie, Jed R Brubaker, and Casey Fiesler. A trade-off-centered framework of content moderation. ACM Transactions on Computer-Human Interaction, 30(1):1–34, 2023.
- [297] Jialun Aaron Jiang, Morgan Klaus Scheuerman, Casey Fiesler, and Jed R Brubaker. Understanding international perceptions of the severity of harmful content online. PloS one, 16(8):e0256762, 2021.
- [298] Isaac Johnson, Connor McMahon, Johannes Schöning, and Brent Hecht. The effect of population and” structural” biases on social media-based algorithms: A case study in geolocation inference across the urban-rural spectrum. In Proceedings of the 2017 CHI conference on Human Factors in Computing Systems, pages 1167–1178, 2017.
- [299] Lindsay Jones. Encyclopedia of Religion: 1. Macmillan Reference USA, 2004.
- [300] Manu Joseph. India faces a linguistic truth: English spoken here. <https://www.nytimes.com/2011/02/17/world/asia/17iht-letter17.html>, 2011.

- [301] Pratik Joshi, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury. The state and fate of linguistic diversity and inclusion in the NLP world. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 6282–6293, Online, July 2020. Association for Computational Linguistics.
- [302] Sanjay Joshi. Contesting histories and nationalist geographies: a comparison of school textbooks in India and Pakistan. South Asian History and Culture, 1(3):357–377, 2010.
- [303] Prerna Juneja and Tanushree Mitra. Auditing e-commerce platforms for algorithmically curated vaccine misinformation. In Proceedings of the 2021 CHI conference on human factors in computing systems, pages 1–27, 2021.
- [304] Prerna Juneja, Deepika Rama Subramanian, and Tanushree Mitra. Through the looking glass: Study of transparency in Reddit’s moderation practices. Proceedings of the ACM on Human-Computer Interaction, 4(GROUP), 2020.
- [305] Jeffrey H Kahn, Renee M Tobin, Audra E Massey, and Jennifer A Anderson. Measuring emotional expression with the linguistic inquiry and word count. The American journal of psychology, 120(2):263–286, 2007.
- [306] R Kalota. A history of the education of the Shudra untouchables before and under the British rule in India, circ. 2000 BC to 1947. PhD thesis, Durham University, 1948.
- [307] Dip Kapoor. Human rights as paradox and equivocation in contexts of adivasi (original dweller) dispossession in India. Journal of Asian and African Studies, 47(4):404–420, 2012.
- [308] Lamia Karim. Pushed to the margins: Adivasi peoples in Bangladesh and the case of kalpana chakma. Contemporary South Asia, 7(3):301–316, 1998.
- [309] Nadia Karizat, Dan Delmonaco, Motahhare Eslami, and Nazanin Andalibi. Algorithmic folk theories and identity: How tiktok users co-produce knowledge of identity and engage in algorithmic resistance. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–44, 2021.
- [310] Sudipta Kaviraj. The making of a language of patriotism in modern bengali. Nationalism and Ethnosymbolism: History, Culture and Ethnicity in the Formation of Nations, pages 248–264, 2007.
- [311] Shirin Keen. Partition of India. <https://scholarblogs.emory.edu/postcolonialstudies/2014/06/21/partition-of-india/>, 1998. Revised: October 2017; Last accessed: May 29, 2020.
- [312] Robin DG Kelley. A poetics of anticolonialism. Monthly Review, 51(6):1, 1999.
- [313] Simon Kemp. Digital 2023: Bangladesh — DataReportal – Global Digital Insights — datareportal.com. <https://datareportal.com/reports/digital-2023-bangladesh>, 2023. [Accessed 12-Jul-2023].
- [314] Mehtab Khan and Alex Hanna. The subjects and stages of ai dataset development: A framework for dataset accountability. Ohio St. Tech. LJ, 19:171, 2022.

- [315] Hissam Khandker. Which India is claiming to have been colonised? <https://www.thedailystar.net/op-ed/politics/which-india-claiming-have-been-colonised-119284>, July 2015. Last accessed: August 4, 2020.
- [316] Shafkat Kibria, Ahnaf Mozib Samin, M Humayon Kobir, M Shahidur Rahman, M Reza Selim, and M Zafar Iqbal. Bangladeshi bangla speech corpus for automatic speech recognition research. Speech Communication, 136, 2022.
- [317] Juho Kim, Eun-Young Ko, Jonghyuk Jung, Chang Won Lee, Nam Wook Kim, and Jihee Kim. Factful: Engaging taxpayers in the public discussion of a government budget. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 2843–2852, 2015.
- [318] Nam Wook Kim, Jonghyuk Jung, Eun-Young Ko, Songyi Han, Chang Won Lee, Juho Kim, and Jihee Kim. Budgetmap: Engaging taxpayers in the issue-driven classification of a government budget. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing, pages 1028–1039, 2016.
- [319] Svetlana Kiritchenko and Saif Mohammad. Examining gender and race bias in two hundred sentiment analysis systems. In Proceedings of the Seventh Joint Conference on Lexical and Computational Semantics, pages 43–53, New Orleans, Louisiana, June 2018. Association for Computational Linguistics.
- [320] Shamika Klassen, Sara Kingsley, Kalyn McCall, Joy Weinberg, and Casey Fiesler. More than a modern day green book: Exploring the online community of black twitter. Proc. of the ACM on Human-Computer Interaction, 5, 2021.
- [321] Kate Klonick. The new governors: The people, rules, and processes governing online speech. Harv. L. Rev., 131:1598, 2017.
- [322] Bart P Knijnenburg, Saadhika Sivakumar, and Daricia Wilkinson. Recommender systems for self-actualization. In Proceedings of the 10th acm conference on recommender systems, pages 11–14, 2016.
- [323] Ramaravind Kommiya Mothilal, Dibyendu Mishra, Sachita Nishal, Faisal M Lalani, and Joyojeet Pal. Voting with the stars: Analyzing partisan engagement between celebrities and politicians in India. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1):1–29, 2022.
- [324] Bill Kovach and Tom Rosenstiel. The elements of journalism, revised and updated 4th edition: What newspeople should know and the public should expect. Crown, 2021.
- [325] Margaret Kovach. Indigenous methodologies: Characteristics, conversations, and contexts. University of Toronto Press, 2010.
- [326] Debbie Kralik, Kate Visentin, and Antonia Van Loon. Transition: a literature review. Journal of advanced nursing, 55(3):320–329, 2006.
- [327] William H Kruskal and W Allen Wallis. Use of ranks in one-criterion variance analysis. Journal of the American statistical Association, 47(260):583–621, 1952.

- [328] Anand Kumar. Impact of west bengal politics on India–Bangladesh relations. Strategic Analysis, 37(3):338–352, 2013.
- [329] Neha Kumar, Nassim Jafarinaimi, and Mehrab Bin Morshed. Uber in Bangladesh: The tangled web of mobility and justice. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–21, 2018.
- [330] Poka Laenui. Processes of decolonization. Reclaiming Indigenous voice and vision, pages 150–160, 2000.
- [331] Marie Lall. Educate to hate: The use of education in the creation of antagonistic national identities in India and pakistan. Compare, 38(1):103–119, 2008.
- [332] Nicolas LaLone. Values levers and the unintended consequences of design. In Proceedings of the companion publication of the 17th ACM conference on Computer supported cooperative work & social computing, pages 189–192, 2014.
- [333] Cliff Lampe and Paul Resnick. Slash (dot) and burn: distributed moderation in a large online conversation space. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 543–550, 2004.
- [334] Cliff AC Lampe, Erik Johnston, and Paul Resnick. Follow the reader: filtering comments on slashdot. In Proceedings of the SIGCHI conference on Human factors in computing systems, pages 1253–1262, 2007.
- [335] James Lane. The 10 most spoken languages in the world. <https://www.babbel.com/en/magazine/the-10-most-spoken-languages-in-the-world>, 2023. [Accessed 23-08-2023].
- [336] Patricia G Lange. Publicly private and privately public: Social networking on youtube. Journal of computer-mediated communication, 13(1):361–380, 2007.
- [337] Shaimaa Lazem, Danilo Giglito, Makuochi Samuel Nkwo, Hafeni Mthoko, Jessica Upani, and Anicia Peters. Challenges and paradoxes in decolonising hci: A critical discussion. Computer Supported Cooperative Work (CSCW), pages 1–38, 2021.
- [338] Christopher A Le Dantec and W Keith Edwards. Designs on dignity: perceptions of technology among the homeless. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 627–636, 2008.
- [339] Débora De Castro Leal, Max Krüger, Vanessa Teles E Teles, Carlos Antônio Teles E Teles, Denise Machado Cardoso, Dave Randall, and Volker Wulf. Digital technology at the edge of capitalism: Experiences from the brazilian amazon rainforest. ACM Transactions on Computer-Human Interaction (TOCHI), 28(3):1–39, 2021.
- [340] Yu-Hao Lee and Gary Hsieh. Does slacktivism hurt activism? the effects of moral balancing and consistency in online activism. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 811–820, 2013.
- [341] Lawrence Leung. Validity, reliability, and generalizability in qualitative research. Journal of family medicine and primary care, 4(3):324, 2015.

- [342] Bryan Li, Dimitrios Dimitriadis, and Andreas Stolcke. Acoustic and lexical sentiment analysis for customer service calls. In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 5876–5880. IEEE, 2019.
- [343] Calvin A Liang, Sean A Munson, and Julie A Kientz. Embracing four tensions in human-computer interaction research with marginalized people. ACM Transactions on Computer-Human Interaction (TOCHI), 28(2):1–47, 2021.
- [344] Michael K Lindell. Disaster studies. Current Sociology, 61(5-6):797–825, 2013.
- [345] Silvia Lindtner, Shaowen Bardzell, and Jeffrey Bardzell. Reconstituting the utopian vision of making: Hci after technosolutionism. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pages 1390–1402, 2016.
- [346] Silvia Lindtner, Shaowen Bardzell, and Jeffrey Bardzell. Design and intervention in the age of” no alternative”. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–21, 2018.
- [347] Sebastian Linxen, Christian Sturm, Florian Brühlmann, Vincent Cassau, Klaus Opwis, and Katharina Reinecke. How weird is chi? In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2021.
- [348] Randall M Livingstone. Make a difference in a different way: Twitter bot creators and wikipedia transparency. Computer Supported Cooperative Work (CSCW), 30(5):733–756, 2021.
- [349] Chris Lloyd, Su Ren Wong, and Leon Petchkovsky. Art and recovery in mental health: A qualitative investigation. British Journal of Occupational Therapy, 70(5):207–214, 2007.
- [350] Ramon Lobato. The cultural logic of digital intermediaries: Youtube multichannel networks. Convergence, 22(4):348–360, 2016.
- [351] Leah Lomoki Lomotey-Nakon. Du bois’s decolonial pragmatism: Teaching community psychology toward epistemological liberation. American journal of community psychology, 62(3-4):364–373, 2018.
- [352] Ania Loomba. Colonialism/postcolonialism. Routledge, 2015.
- [353] Peter Loshin. Resource description framework (rdf). <https://www.techtarget.com/searcharchitecture/definition/Resource-Description-Framework-RDF>, 2022. Last accessed: February 11, 2023.
- [354] Zhicong Lu, Michelle Annett, Mingming Fan, and Daniel Wigdor. ” i feel it is my responsibility to stream” streaming and engaging with intangible cultural heritage through livestreaming. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2019.
- [355] María Lugones. Heterosexualism and the colonial/modern gender system. Hypatia, 22(1):186–219, 2007.
- [356] Maria Lugones. The coloniality of gender. Springer, 2016.

- [357] Haiwei Ma, C Estelle Smith, Lu He, Saumik Narayanan, Robert A Giaquinto, Roni Evans, Linda Hanson, and Svetlana Yarosh. Write for life: Persisting in online health communities through expressive writing and social support. Proceedings of the ACM on Human-Computer Interaction, 1(CSCW):1–24, 2017.
- [358] Renkai Ma, Xinning Gui, and Yubo Kou. Multi-platform content creation: the configuration of creator ecology through platform prioritization, content synchronization, and audience management. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pages 1–19, 2023.
- [359] Renkai Ma and Yubo Kou. ” how advertiser-friendly is my video?”: Youtuber’s socio-economic interactions with algorithmic content moderation. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2), 2021.
- [360] Triloki Nath Madan. Two faces of bengali ethnicity: Muslim bengali or bengali muslim. The developing economies, 10(1):74–85, 1972.
- [361] Boria Majumdar. Nationalist romance to postcolonial sport: Cricket in 2006 India. Sport in Society, 10(1), 2007.
- [362] Nelson Maldonado-Torres. On the coloniality of being: Contributions to the development of a concept. Cultural studies, 21(2-3):240–270, 2007.
- [363] Henry B Mann and Donald R Whitney. On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics, pages 50–60, 1947.
- [364] Michael Mann. South Asia’s modern history: Thematic perspectives. Routledge, 2014.
- [365] Asma Mansoor and Samina Malik. Decolonial-posthuman pedagogy and conceptualising ‘womanness’ in postcolonial pakistan. Compare: A Journal of Comparative and International Education, 52(3):440–456, 2022.
- [366] Gloria Mark and Bryan Semaan. Resilience in collaboration: Technology as a resource for new patterns of action. In Proceedings of the 2008 ACM conference on Computer supported cooperative work, pages 137–146, 2008.
- [367] Gloria Mark and Bryan Semaan. Expanding a country’s borders during war: the internet war diary. In Proceedings of the 2009 international workshop on Intercultural collaboration, pages 3–12, 2009.
- [368] Gloria J Mark, Ban Al-Ani, and Bryan Semaan. Resilience through technology adoption: merging the old and the new in iraq. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 689–698, 2009.
- [369] Misa Maruyama, Scott P Robertson, Sara Douglas, Roxanne Raine, and Bryan Semaan. Social watching a civic broadcast: Understanding the effects of positive feedback and other users’ opinions. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 794–807, 2017.

- [370] Vadim Mikhaïlovich Masson, János Harmatta, Baij Nath Puri, GF Etemadi, Boris Anatolevich Litvinskiï, Guangda Zhang, R Shabani Samghabadi, Chahryar Adle, Irfan Habib, Anara Tabyshalieva, et al. History of civilizations of Central Asia. UNESCO, 1992.
- [371] Brian McBride. The resource description framework (rdf) and its vocabulary description language rdfs. Handbook on ontologies, pages 51–65, 2004.
- [372] Leslie McCall. The complexity of intersectionality. Signs: Journal of women in culture and society, 30(3):1771–1800, 2005.
- [373] Jo McCormack, Murray Pratt, and Alistair Rolls Alistair Rolls. Hexagonal variations: diversity, plurality and reinvention in contemporary France, volume 359. Rodopi, 2011.
- [374] Nora McDonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–23, 2019.
- [375] Mary L McHugh. The chi-square test of independence. Biochemia medica, 23(2):143–149, 2013.
- [376] John McLeod. The history of India. ABC-CLIO, 2015.
- [377] Sarah McRoberts, Elizabeth Bonsignore, Tamara Peyton, and Svetlana Yarosh. Do it for the viewers! audience engagement behaviors of young youtubers. In Proc. of International Conference on Interaction Design and Children, 2016.
- [378] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. ACM computing surveys (CSUR), 54(6):1–35, 2021.
- [379] Monica Melanchthon. Persecution of Indian christians. Dialog, 41(2):103–113, 2002.
- [380] Sharan B Merriam and Robin S Grenier. Qualitative research in practice: Examples for discussion and analysis. John Wiley & Sons, 2019.
- [381] Samantha Merritt and Shaowen Bardzell. Postcolonial language and culture theory for hci4d. In CHI’11 Extended Abstracts on Human Factors in Computing Systems, pages 1675–1680. ACM New York, NY, USA, 2011.
- [382] Danaë Metaxa, Joon Sung Park, Ronald E Robertson, Karrie Karahalios, Christo Wilson, Jeff Hancock, Christian Sandvig, et al. Auditing algorithms: Understanding algorithmic systems from the outside in. Foundations and Trends® in Human-Computer Interaction, 14(4):272–344, 2021.
- [383] Walter D Mignolo. Delinking: The rhetoric of modernity, the logic of coloniality and the grammar of de-coloniality. Cultural studies, 21(2-3):449–514, 2007.
- [384] Walter D Mignolo. Local histories/global designs. Princeton University Press, 2012.
- [385] Walter D Mignolo. The Politics of Decolonial Investigations. Duke University Press, 2021.
- [386] Walter D Mignolo and Arturo Escobar. Globalization and the decolonial option. Routledge, 2013.

- [387] Walter D Mignolo and Catherine E Walsh. On decoloniality: Concepts, analytics, praxis. Duke University Press, 2018.
- [388] Mary Milliken, Kerri Gibson, and Susan O'Donnell. User-generated video and the online public sphere: Will youtube facilitate digital freedom of expression in atlantic canada. American Communication Journal, 10(3), 2008.
- [389] Mary Milliken, Kerri Gibson, Susan O'Donnell, Janice Singer, et al. User-generated online video and the atlantic canadian public sphere: A youtube study. In Proceedings of the international communication association annual conference, pages 22–26, 2008.
- [390] Mary C Milliken and Susan O'Donnell. User-generated online video: The next public sphere? In 2008 IEEE International Symposium on Technology and Society, pages 1–3. IEEE, 2008.
- [391] Mary C Milliken, Susan O'Donnell, Kerri Gibson, and Betty Daniels. Older citizens and video communications: A case study. The Journal of Community Informatics, 8(1), 2012.
- [392] Nusrat Jahan Mim and Syed Ishtiaque Ahmed. Others' images: Online social media, architectural improvisations, and spatial marginalization in Bangladesh. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–14, 2020.
- [393] Maya Mirchandani. Digital hatred, real violence: Majoritarian radicalisation and social media in India. ORF Occasional Paper, 167:1–30, 2018.
- [394] Dibyendu Mishra, Syeda Zainab Akbar, Arshia Arya, Saloni Dash, Rynaa Grover, and Joyojeet Pal. Rihanna versus bollywood: Twitter influencers and the Indian farmers' protest. arXiv preprint [arXiv:2102.04031](https://arxiv.org/abs/2102.04031), 2021.
- [395] Arpeeta Shams Mizan. Identity crisis and the pseudo-minorities in Bangladesh: Is the right to cultural identity the answer? International Journal on Minority and Group Rights, 29(1):1–32, 2021.
- [396] Arpeeta Shams Mizan and Syed Ishtiaque Ahmed. Silencing the minority through domination in social media platform: Impact on the pluralistic Bangladeshi society. ELCOP Yearbook of Human Rights (2018), 2019.
- [397] Shakir Mohamed, Marie-Therese Png, and William Isaac. Decolonial ai: Decolonial theory as sociotechnical foresight in artificial intelligence. Philosophy & Technology, 33:659–684, 2020.
- [398] Maria D Molina and S Shyam Sundar. When ai moderates online content: effects of human collaboration and interactive transparency on user trust. Journal of Computer-Mediated Communication, 27(4):zmac010, 2022.
- [399] Tsubasa Morioka, Nicole B Ellison, and Michael Brown. Identity work on social media sites: Disadvantaged students' college transition processes. In Proceedings of the 19th ACM conference on computer-supported cooperative work & social computing, pages 848–859, 2016.
- [400] Jonathan W Moses and Torbjørn L Knutsen. Ways of knowing: Competing methodologies in social and political research. Bloomsbury Publishing, 2019.
- [401] Migration Policy International (MPI). Bangladeshi diaspora in the united states. <https://www.migrationpolicy.org/sites/default/files/publications/RAD-{B}angladesh.pdf>, 2014. Last accessed: May 31, 2024.

- [402] Isabel Munoz, Michael Dunn, Steve Sawyer, and Emily Michaels. Platform-mediated markets, online freelance workers and deconstructed identities. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW2), 2022.
- [403] Sarah Myers West. Censored, suspended, shadowbanned: User interpretations of content moderation on social media platforms. New Media & Society, 20(11):4366–4383, 2018.
- [404] Seema Nagar, Sameer Gupta, CS Bahushruth, Ferdous A Barbhuiya, and Kuntal Dey. Empirical assessment and characterization of homophily in classes of hate speeches. In AffCon@AAAI, pages 30–34, 2021.
- [405] Ved P Nanda. Self-determination outside the colonial context: The birth of Bangladesh in retrospect. In Self-determination, pages 193–220. Routledge, 2019.
- [406] Ashis Nandy. The Intimate Enemy: Loss and Recovery of Self Under Colonialism. Oxford University Press Oxford, 1989.
- [407] Sabelo J Ndlovu-Gatsheni. Epistemic freedom in Africa: Deprovincialization and decolonization. Routledge, 2018.
- [408] Matti Nelimarkka, Jean Philippe Rancy, Jennifer Grygiel, and Bryan Semaan. (re) design to mitigate political polarization: Reflecting habermas’ ideal communication space in the united states of america and finland. Proceedings of the ACM on Human-computer Interaction, 3(CSCW):1–25, 2019.
- [409] BBC News. Article 370: What happened with kashmir and why it matters. <https://www.bbc.com/news/world-asia-india-49234708>, 2019. Last accessed: October 1, 2020.
- [410] BBC News. Citizenship amendment bill: India’s new ‘anti-muslim’ law explained. <https://www.bbc.com/news/world-asia-india-50670393>, 2019. Last accessed: October 1, 2020.
- [411] Gabriel Nicholas and Aliya Bhatia. Lost in translation: Large language models in non-english content analysis. arXiv preprint arXiv:2306.07377, 2023.
- [412] Gabriel Nicholas and Aliya Bhatia. Re: Request for information (rfi) on developing a roadmap for the directorate for technology, innovation, and partnerships at the national science foundation| 88 fr 26345. 2023.
- [413] Helen Nissenbaum. How computer systems embody values. Computer, 34(3):120–119, 2001.
- [414] Jim Norwine and Alfonso Gonzalez. The Third World: States of mind and being. Taylor & Francis, 1988.
- [415] Fayika Farhat Nova, Michael Ann DeVito, Pratyasha Saha, Kazi Shohanur Rashid, Shashwata Roy Turzo, Sadia Afrin, and Shion Guha. ” facebook promotes more harassment” social media ecosystem, skill and marginalized hijra identity in Bangladesh. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW1):1–35, 2021.
- [416] Nicole Novielli, Fabio Calefato, and Filippo Lanubile. A gold standard for emotion annotation in stack overflow. In 2018 IEEE/ACM 15th International Conference on Mining Software Repositories (MSR), pages 14–17. IEEE, 2018.

- [417] Vera Nünning and Ansgar Nünning. Fictions of empire and the (un-) making of imperialist mentalities: colonial discourse and post-colonial criticism revisited. FOREIGN LITERATURE STUDIES, 37(4):22–38, 2015.
- [418] Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464):447–453, 2019.
- [419] Nigini Oliveira, Michael Muller, Nazareno Andrade, and Katharina Reinecke. The exchange in stackexchange: Divergences between stack overflow and its culturally diverse participants. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–22, 2018.
- [420] Alexandra Olteanu, Carlos Castillo, Jeremy Boy, and Kush Varshney. The effect of extremist violence on hateful speech online. In Proceedings of the international AAAI conference on web and social media, volume 12/1, 2018.
- [421] Gail Omvedt. Dalits and the democratic revolution: Dr Ambedkar and the Dalit movement in colonial India. SAGE, 1994.
- [422] James Onley. The raj reconsidered: British India’s informal empire and spheres of influence in asia and africa. Asian Affairs, 40(1):44–62, 2009.
- [423] Carsten Østerlund, Kevin Crowston, and Corey Jackson. Building an apparatus: Refractive, reflective, and diffractive readings of trace data. Journal of the Association for Information Systems, 21(1):10, 2020.
- [424] Marie Ozanne, Aparajita Bhandari, Natalya Bazarova, and Dominic DiFranzo. Shall ai moderators be made visible? perception of accountability and trust in moderation systems on social media platforms. Big Data & Society, 2022.
- [425] Leysia Palen and Amanda L Hughes. Social media in disaster communication. In Handbook of disaster research, pages 497–518. Springer, 2018.
- [426] Leysia Palen, Sarah Vieweg, and Kenneth Mark Anderson. Supporting “everyday analysts” in safety-and time-critical situations. The Information Society, 27(1):52–62, 2011.
- [427] Christina A Pan, Sahil Yakhmi, Tara P Iyer, Evan Strasnick, Amy X Zhang, and Michael S Bernstein. Comparing the perceived legitimacy of content moderation processes: Contractors, algorithms, expert panels, and digital juries. Proceedings of the ACM on Human-Computer Interaction, 6(CSCW1):1–31, 2022.
- [428] Anmol Panda, Sunandan Chakraborty, Noopur Raval, Han Zhang, Mugdha Mohapatra, Syeda Zainab Akbar, and Joyojeet Pal. Affording extremes: incivility, social media and democracy in the Indian context. In Proceedings of the 2020 International Conference on Information and Communication Technologies and Development, pages 1–12, 2020.
- [429] Anmol Panda, Ramaravind Kommiya Mothilal, Monojit Choudhury, Kalika Bali, and Joyojeet Pal. Topical focus of political campaigns and its impact: Findings from politicians’ hashtag use during the 2019 Indian elections. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1):1–14, 2020.

- [430] Gyanendra Pandey. In defense of the fragment: writing about Hindu-Muslim riots in India today. Representations, 37:27–55, 1992.
- [431] Gyanendra Pandey et al. Remembering partition: Violence, nationalism and history in India, volume 7. Cambridge University Press, 2001.
- [432] Bhasa Vidya Parishad. Praci Bhasavijnan: Indian Journal of Linguistics, volume 20. Bhasa Vidya Parishad., 2001.
- [433] Prasannan Parthasarathi. Why Europe grew rich and Asia did not: Global economic divergence, 1600–1850. Cambridge University Press, 2011.
- [434] Sujata Patel. Sociological study of religion: Colonial modernity and 19th century majoritarianism. Economic and Political Weekly, pages 1089–1094, 2007.
- [435] David Patterson. How we’re minimizing ai’s carbon footprint. <https://blog.google/technology/ai/minimizing-carbon-footprint/>, 2021. Last accessed: August 3, 2023.
- [436] Umashanthi Pavalanathan, Xiaochuang Han, and Jacob Eisenstein. Mind your pov: Convergence of articles and editors towards wikipedia’s neutrality norm. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–23, 2018.
- [437] Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50(302):157–175, 1900.
- [438] James W Pennebaker. Writing about emotional experiences as a therapeutic process. Psychological science, 8(3):162–166, 1997.
- [439] Maayan Perel and Niva Elkin-Koren. Accountability in algorithmic copyright enforcement. Stan. Tech. L. Rev., 19:473, 2015.
- [440] Kavita Philip, Lilly Irani, and Paul Dourish. Postcolonial computing: A tactical survey. Science, Technology, & Human Values, 37(1):3–29, 2012.
- [441] Alix Philippon. 11w an ambiguous and contentious politicization of sufi shrines and pilgrimages in pakistan. Devotional Islam in Contemporary South Asia: Shrines, Journeys and Wanderers, page 174, 2015.
- [442] Nathaniel Poor. Mechanisms of an online public sphere: The website slashdot. Journal of computer-mediated communication, 10(2):JCMC1028, 2005.
- [443] Vinodkumar Prabhakaran, Ben Hutchinson, and Margaret Mitchell. Perturbation sensitivity analysis to detect unintended model biases. arXiv preprint arXiv:1910.04210, 2019.
- [444] Anibal Quijano. Coloniality of power and eurocentrism in latin america. International Sociology, 15(2):215–232, 2000.
- [445] Natasha Raheja. Neither here nor there: Pakistani Hindu refugee claims at the interface of the international and South Asian refugee regimes. Journal of Refugee Studies, 31(3):334–352, 2018.

- [446] Aziz Rahman, Mohsin Ali, and Saad Kahn. The british art of colonialism in India: Subjugation and division. Peace and Conflict Studies, 25(1):5, 2018.
- [447] Inioluwa Deborah Raji, I Elizabeth Kumar, Aaron Horowitz, and Andrew Selbst. The fallacy of ai functionality. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, pages 959–972, 2022.
- [448] Maia Ramnath. Decolonizing anarchism: an antiauthoritarian history of India’s liberation struggle. AK Press, 2012.
- [449] Gavin Rand. ‘martial races’ and ‘imperial subjects’: violence and governance in colonial India, 1857–1914. European Review of History: Revue européenne d’histoire, 13(1):1–20, 2006.
- [450] Mohammad Rifat Ahmmad Rashid, Kazi Ferdous Hasan, Md. Rakibul Hasan, Aritra Das, and Mithila Sultana. A multilabel multiclass sentiment and emotion dataset from Bangladeshi e-commerce reviews. <https://data.mendeley.com/datasets/rzjfg7t9kf/>, 2023. Last accessed: Dec 25, 2023.
- [451] MN Ravishankar, Shan L Pan, and Michael D Myers. Information technology offshoring in India: a postcolonial perspective. European Journal of Information Systems, 22(4):387–402, 2013.
- [452] Elizabeth Reid. Hierarchy and power. Communities in cyberspace, pages 107–133, 1999.
- [453] Ernest Renan. What is a nation? And other political writings. Columbia University Press, 2018.
- [454] UW Copyright Resources. Digital millennium copyright act (dmca). <https://copyrightresource.uw.edu/copyright-law/dmca/>, n.d. Last accessed: June 23, 2024.
- [455] Nani Jansen Reventlow. How artificial intelligence impacts marginalised groups. <https://digitalfreedomfund.org/how-artificial-intelligence-impacts-marginalised-groups/>, 2021. Last accessed: Sep 11, 2023.
- [456] World Population Review. World population by country. <https://worldpopulationreview.com/>, 2021. [Accessed: July 2, 2021].
- [457] Ali Riaz. God willing: The politics of Islamism in Bangladesh. Rowman & Littlefield, 2004.
- [458] René Riedl. Is trust in artificial intelligence systems related to user personality? review of empirical evidence and future research directions. Electronic Markets, 32(4):2021–2051, 2022.
- [459] Mohammad Rashidujjaman Rifat, Dipto Das, Arpon Podder, Mahiratul Jannat, Robert Soden, Bryan Semaan, and Syed Ishtiaque Ahmed. The politics of fear and the experience of Bangladeshi religious minority communities using social media platforms. Proceedings of the ACM on human-computer interaction, (CSCW):1–31, 2024. in press.
- [460] Mohammad Rashidujjaman Rifat, Mahiratul Jannat, Mahdi Nasrullah Al-Ameen, SM Taibatul Haque, Muhammad Ashad Kabir, and Syed Ishtiaque Ahmed. Purdah, amanah, and gheebat: Understanding privacy in Bangladeshi “pious” Muslim communities. In Proceedings of the 4th ACM SIGCAS Conference on Computing and Sustainable Societies, pages 199–214, 2021.

- [461] Mohammad Rashidujjaman Rifat, Toha Toriq, and Syed Ishtiaque Ahmed. Religion and sustainability: Lessons of sustainable computing from islamic religious communities. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2):1–32, 2020.
- [462] Herbert Hope Risley. The study of ethnology in India. The Journal of the Anthropological Institute of Great Britain and Ireland, 20:235–263, 1891.
- [463] Jeffrey A Roberts, Il-Horn Hann, and Sandra A Slaughter. Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the apache projects. Management science, 52(7):984–999, 2006.
- [464] Sarah T Roberts. Commercial content moderation: Digital laborers’ dirty work. Media Studies Publications, 12, 2016.
- [465] Ronald E Robertson, Shan Jiang, Kenneth Joseph, Lisa Friedland, David Lazer, and Christo Wilson. Auditing partisan audience bias within google search. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 2018.
- [466] Julia Romberg and Tobias Escher. Making sense of citizens’ input through artificial intelligence: A review of methods for computational text analysis to support the evaluation of contributions in public participation. Digital Government: Research and Practice, 2023.
- [467] Daniela Rosner. Whose stories underpin design? reworking our methods with design fabulations. Interactions, 26(5):20–21, 2019.
- [468] Tirthankar Roy. Where is Bengal?: Situating an Indian region in the early modern world economy. Past & present, 213(1):115–146, 2011.
- [469] Minna Ruckenstein and Linda Lisa Maria Turunen. Re-humanizing the platform: Content moderators and the logic of care. new media & society, 22(6):1026–1042, 2020.
- [470] Christian Rudder. Inside okcupid: The math of online dating. <https://ed.ted.com/lessons/inside-okcupid-the-math-of-online-dating-christian-rudder>, 2013. [Accessed 20-08-2023].
- [471] Henrik Skaug Sætra. Ai in context and the sustainable development goals: Factoring in the unsustainability of the sociotechnical system. Sustainability, 13(4):1738, 2021.
- [472] Niranjana Sahoo. Mounting majoritarianism and political polarization in India. Political Polarization in South and Southeast Asia, 2020.
- [473] Edward W Said. Orientalism. Vintage, 1979.
- [474] Flora Sakketou, Joan Plepi, Riccardo Cervero, Henri Jacques Geiss, Paolo Rosso, and Lucie Flek. FACTOID: A new dataset for identifying misinformation spreaders and political bias. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 3231–3241, Marseille, France, June 2022. European Language Resources Association.
- [475] Christian Sandvig, Kevin Hamilton, Karrie Karahalios, and Cedric Langbort. Auditing algorithms: Research methods for detecting discrimination on internet platforms. Data and discrimination: converting critical concerns into productive inquiry, 22(2014):4349–4357, 2014.

- [476] Maarten Sap, Dallas Card, Saadia Gabriel, Yejin Choi, and Noah A Smith. The risk of racial bias in hate speech detection. In Proceedings of the 57th annual meeting of the association for computational linguistics, pages 1668–1678, 2019.
- [477] Sumit Sarkar. The Swadeshi Movement in Bengal, 1903-1908. People’s Publishing House New Delhi, 1973.
- [478] Shahadat Hossain Sarker. India-Bangladesh border killing: Implication on bilateral relation. 2015.
- [479] Beatrice Savoldi, Marco Gaido, Luisa Bentivogli, Matteo Negri, and Marco Turchi. Under the morphosyntactic lens: A multifaceted evaluation of gender bias in speech translation. In Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 1807–1824, Dublin, Ireland, May 2022. Association for Computational Linguistics.
- [480] Steve Sawyer and Mohammad Hossein Jarrahi. Sociotechnical approaches to the study of information systems. In Computing handbook, third edition: Information systems and information technology, pages 5–1. CRC Press, 2014.
- [481] Devansh Saxena, Erina Seh-Young Moon, Aryan Chaurasia, Yixin Guan, and Shion Guha. Rethinking” risk” in algorithmic systems through a computational narrative analysis of casenotes in child-welfare. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pages 1–19, 2023.
- [482] Devansh Saxena, Seh Young Moon, Dahlia Shehata, and Shion Guha. Unpacking invisible work practices, constraints, and latent power relationships in child welfare through casenote analysis. In Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems, pages 1–22, 2022.
- [483] Devansh Saxena, Charles Repaci, Melanie D Sage, and Shion Guha. How to train a (bad) algorithmic caseworker: A quantitative deconstruction of risk assessments in child welfare. In CHI Conference on Human Factors in Computing Systems Extended Abstracts, pages 1–7, 2022.
- [484] Salim Sazed. Cross-lingual sentiment classification in low-resource bengali language. In Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020), pages 50–60, 2020.
- [485] Fabian Schäfer, Stefan Evert, and Philipp Heinrich. Japan’s 2014 general election: Political bots, right-wing internet activism, and prime minister shinzō abe’s hidden nationalist agenda. Big data, 5(4):294–309, 2017.
- [486] Morgan Klaus Scheuerman, Alex Hanna, and Emily Denton. Do datasets have politics? disciplinary values in computer vision dataset development. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–37, 2021.
- [487] Morgan Klaus Scheuerman, Madeleine Pape, and Alex Hanna. Auto-essentialization: Gender in automated facial analysis as extended colonial project. Big Data & Society, 8(2):20539517211053712, 2021.

- [488] Morgan Klaus Scheuerman, Jacob M Paul, and Jed R Brubaker. How computers see gender: An evaluation of gender classification in commercial facial analysis services. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–33, 2019.
- [489] Morgan Klaus Scheuerman, Kandrea Wade, Caitlin Lustig, and Jed R Brubaker. How we’ve taught algorithms to see identity: Constructing race and gender in image databases for facial analysis. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1):1–35, 2020.
- [490] Morgan Klaus Scheuerman, Katy Weathington, Tarun Mugunthan, Emily Denton, and Casey Fiesler. From human to data to dataset: Mapping the traceability of human subjects in computer vision datasets. Proceedings of the ACM on Human-Computer Interaction, 7(CSCW1):1–33, 2023.
- [491] Ari Schlesinger, W Keith Edwards, and Rebecca E Grinter. Intersectional hci: Engaging identity through gender, race, and class. In Proceedings of the 2017 CHI conference on human factors in computing systems, pages 5412–5427, 2017.
- [492] Victoria Schofield. Kashmir in conflict: India, Pakistan and the unending war. Bloomsbury Publishing, 2021.
- [493] Howard Schuman. A note on the rapid rise of mass bengali nationalism in east pakistan. American Journal of Sociology, 78(2):290–298, 1972.
- [494] Kailah Sebastian. Distinguishing between the strains grounded theory: Classical, interpretive and constructivist. Journal for Social Thought, 3(1), 2019.
- [495] Joseph Seering, Geoff Kaufman, and Stevie Chancellor. Metaphors in moderation. New Media & Society, page 1461444820964968, 2020.
- [496] Joseph Seering, Robert Kraut, and Laura Dabbish. Shaping pro and anti-social behavior on twitch through moderation and example-setting. In Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing, pages 111–125, 2017.
- [497] Joseph Seering, Tony Wang, Jina Yoon, and Geoff Kaufman. Moderator engagement and community development in the age of algorithms. New Media & Society, 21(7):1417–1443, 2019.
- [498] Bryan Semaan. ‘routine infrastructuring’as’ building everyday resilience with technology’ when disruption becomes ordinary. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–24, 2019.
- [499] Bryan Semaan, Lauren M Britton, and Bryan Dosono. Military masculinity and the travails of transitioning: Disclosure in social media. In Proceedings of the 2017 ACM Conference on computer supported cooperative work and social computing, pages 387–403, 2017.
- [500] Bryan Semaan, Bryan Dosono, and Lauren M Britton. Impression management in high context societies: ‘saving face’with ict. In Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing, pages 712–725, 2017.
- [501] Bryan Semaan, Heather Faucett, Scott P Robertson, Misa Maruyama, and Sara Douglas. Designing political deliberation environments to support interactions in the public sphere. In Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems, pages 3167–3176, 2015.

- [502] Bryan Semaan and Gloria Mark. Creating a context of trust with icts: restoring a sense of normalcy in the environment. In Proceedings of the ACM 2011 conference on Computer supported cooperative work, pages 255–264, 2011.
- [503] Bryan Semaan and Gloria Mark. Technology-mediated social arrangements to resolve breakdowns in infrastructure during ongoing disruption. ACM Transactions on Computer-Human Interaction (TOCHI), 18(4):1–21, 2011.
- [504] Bryan C Semaan, Lauren M Britton, and Bryan Dosono. Transition resilience with icts: 'identity awareness' in veteran re-integration. In Proceedings of the 2016 CHI Conference on human factors in computing systems, pages 2882–2894, 2016.
- [505] Bryan C Semaan, Scott P Robertson, Sara Douglas, and Misa Maruyama. Social media supporting political deliberation across multiple public spheres: towards depolarization. In Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing, pages 1409–1421, 2014.
- [506] Dwaipayan Sen. The decline of the caste question: Jogendranath Mandal and the defeat of Dalit politics in Bengal. Cambridge University Press, 2018.
- [507] Shilad Sen, Margaret E Giesel, Rebecca Gold, Benjamin Hillmann, Matt Lesicko, Samuel Naden, Jesse Russell, Zixiao Wang, and Brent Hecht. "Turkers, scholars," "arafat" and "peace" cultural communities and algorithmic gold standards. In Proceedings of the 18th acm conference on computer supported cooperative work & social computing, 2015.
- [508] Agrima Seth, Soham De, Arshia Arya, Steven Wilkinson, Sushant Singh, and Joyojeet Pal. Closed ranks: The discursive value of military support for Indian politicians on social media. In Proceedings of the 2022 International Conference on Information and Communication Technologies and Development, pages 1–11, 2022.
- [509] Ahmed Shafiqul Huque. The impact of colonialism: thoughts on politics and governance in Bangladesh. Asian Affairs, 28(1):15–27, 1997.
- [510] P Ravi Shankar and Palaian Subish. Fair skin in south asia: an obsession? Journal of Pakistan Association of Dermatology, 17(2):100–104, 2016.
- [511] Samuel Sanford Shapiro and Martin B Wilk. An analysis of variance test for normality (complete samples). Biometrika, 52(3/4):591–611, 1965.
- [512] Eva Sharma, Koustuv Saha, Sindhu Kiranmai Ernala, Sucheta Ghoshal, and Munmun De Choudhury. Analyzing ideological discourse on social media: A case study of the abortion debate. In Proceedings of the 2017 international conference of the computational social science society of the americas, pages 1–8, 2017.
- [513] Wendy S Shaw, R Douglas K Herman, and G Rebecca Dobbs. Encountering indigeneity: Re-imagining and decolonizing geography. Geografiska Annaler: Series B, Human Geography, 88(3):267–276, 2006.
- [514] Natasha Shevde. All's fair in love and cream: A cultural case study of fair & lovely in India. Advertising & society review, 9(2), 2008.

- [515] Yongren Shi, Kai Mast, Ingmar Weber, Agrippa Kellum, and Michael Macy. Cultural fault lines and political polarization. In Proceedings of the 2017 ACM on web science conference, pages 213–217, 2017.
- [516] Irina Shklovski, Robert Kraut, and Jonathon Cummings. Keeping in touch by technology: Maintaining friendships after a residential move. In Proceedings of the sigchi conference on human factors in computing systems, pages 807–816, 2008.
- [517] Michael L Siciliano. Intermediaries in the age of platformized gatekeeping: The case of youtube “creators” and mens in the us. Poetics, 97:101748, 2023.
- [518] Ellen Simpson and Bryan Semaan. For you, or for” you”? everyday lgbtq+ encounters with tiktok. Proceedings of the ACM on human-computer interaction, 4(CSCW3):1–34, 2021.
- [519] Ellen Simpson and Bryan Semaan. Rethinking creative labor: A sociotechnical examination of creativity & creative work on tiktok. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, pages 1–16, 2023.
- [520] Philipp Singer, Fabian Flöck, Clemens Meinhart, Elias Zeitfogel, and Markus Strohmaier. Evolution of Reddit: from the front page of the internet to a self-referential community? In Proceedings of the 23rd international conference on world wide web, pages 517–522, 2014.
- [521] Gurharpal Singh and Heewon Kim. The limits of India’s ethno-linguistic federation: Understanding the demise of sikh nationalism. Regional & Federal Studies, 28(4):427–445, 2018.
- [522] Munindar P Singh. Norms as a basis for governing sociotechnical systems. ACM Transactions on Intelligent Systems and Technology (TIST), 5(1):1–23, 2014.
- [523] Manjira Sinha and Anupam Basu. A study of readability of texts in bangla through machine learning approaches. Education and information technologies, 21(5):1071–1094, 2016.
- [524] Mrinalini Sinha. Colonial masculinity: The ‘manly englishman’and the ‘effeminate bengali’in the late nineteenth century. In Colonial masculinity. Manchester University Press, 2017.
- [525] A Duval Smith. Problems of conflict management in virtual communities. Communities in cyberspace, pages 134–163, 1999.
- [526] Andrea Smith. Not an Indian tradition: The sexual colonization of native peoples. Hypatia, 18(2):70–85, 2003.
- [527] Hayden Smith and William Cipolli. The instagram/facebook ban on graphic self-harm imagery: A sentiment analysis and topic modeling approach. Policy & Internet, 14(1):170–185, 2022.
- [528] Linda Tuhiwai Smith. Decolonizing methodologies: Research and indigenous peoples. Bloomsbury Publishing, 2021.
- [529] Madeline E Smith, Duyen T Nguyen, Charles Lai, Gilly Leshed, and Eric PS Baumer. Going to college and staying connected: Communication between college freshmen and their parents. In Proceedings of the ACM 2012 conference on computer supported cooperative work, pages 789–798, 2012.

- [530] David Snow. Collective identity and expressive forms. International Encyclopedia of the Social & Behavioral Sciences, pages 2212–2219, 2001.
- [531] David A Snow and Leon Anderson. Identity work among the homeless: The verbal construction and avowal of personal identities. American journal of sociology, 92(6):1336–1371, 1987.
- [532] Richard Socher, Alex Perelygin, Jean Wu, Jason Chuang, Christopher D Manning, Andrew Y Ng, and Christopher Potts. Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 conference on empirical methods in natural language processing, pages 1631–1642, 2013.
- [533] Robert Soden and Leysia Palen. Informating crisis: Expanding critical perspectives in crisis informatics. Proceedings of the ACM on human-computer interaction, 2(CSCW):1–22, 2018.
- [534] Robert Soden, David Ribes, Seyram Avle, and Will Sutherland. Time for historicism in cscw: An invitation. Proceedings of the ACM on Human-Computer Interaction, 5(CSCW2):1–18, 2021.
- [535] Anders Søgaard, Anders Johannsen, Barbara Plank, Dirk Hovy, and Héctor Martínez Alonso. What’s in a p-value in nlp? In Proceedings of the eighteenth conference on computational natural language learning, pages 1–10, 2014.
- [536] Leron Solomon. Fair users or content abusers: The automatic flagging of non-infringing videos by content id on youtube. Hofstra L. Rev., 44:237, 2015.
- [537] Gayatri Chakravorty Spivak. Can the subaltern speak? Die Philosophin, 14(27):42–58, 2003.
- [538] Ramya Srinivasan and Kanji Uchino. Biases in generative art: A causal look from the lens of art history. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, pages 41–51, 2021.
- [539] Nick Srnicek. Platform capitalism. John Wiley & Sons, 2017.
- [540] The Daily Star. Cricketer liton das’ puja greetings draw hate comments. <https://www.thedailystar.net/sports/cricketer-liton-das%E2%80%99puja-greetings-draw-hate-comments-160513>, 2015. Last accessed: Jan 1, 2023.
- [541] Kate Starbird, Ahmer Arif, and Tom Wilson. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), 2019.
- [542] Tripura State Portal. Know tripura. <https://web.archive.org/web/20150212025154/http://tripura.gov.in/knowtripura>, 2015. Last accessed: December 27, 2020.
- [543] Statista. India: Muslim population share by state. <https://www.statista.com/statistics/1317643/india-muslim-population-share-by-state/>, 2011. [Accessed 15-Jul-2023].
- [544] Statista. YouTube users by country 2023 | Statista — statista.com. <https://www.statista.com/statistics/280685/number-of-monthly-unique-youtube-users/>, 2023. [Accessed 12-Jul-2023].

- [545] Neil Stewart. Divide and rule: British policy in indian history. Science & Society, pages 49–57, 1951.
- [546] Ann Laura Stoler. Carnal knowledge and imperial power: Race and the intimate in colonial rule. Univ of California Press, 2010.
- [547] Anselm Strauss and Juliet Corbin. Grounded theory methodology. Handbook of qualitative research, 17(1):273–285, 1994.
- [548] Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. arXiv preprint arXiv:1906.02243, 2019.
- [549] Student. The probable error of a mean. Biometrika, 6(1):1–25, 1908.
- [550] Carl Franklin Stychin. A nation by rights: National cultures, sexual identity politics, and the discourse of rights, volume 16. Temple University Press, 1998.
- [551] Christopher Su. pyquora. <https://github.com/csu/pyquora>, 2016. PyPI: <https://pypi.org/project/quora/>, last accessed: June 18, 2020.
- [552] Tanka B Subba. Indigenising the limbus: Trajectory of a nation divided into two nationstates. Indigeneity in India, pages 143–158, 2006.
- [553] Heng Sun and Wan Ni. Design and application of an ai-based text content moderation system. Scientific Programming, 2022, 2022.
- [554] Jayaraj Sundaresan. Decolonial reflections on urban pedagogy in India. Area, 52(4):722–730, 2020.
- [555] Harsh Suri. Purposeful sampling in qualitative research synthesis. Qualitative research journal, 11(2):63–75, 2011.
- [556] Matt Swayne. Users trust ai as much as humans for flagging problematic content | penn state university. <https://www.psu.edu/news/institute-computational-and-data-sciences/story/users-trust-ai-much-humans-flagging-problematic/>, 2022. Last accessed: August 10, 2023.
- [557] Latanya Sweeney. Discrimination in online ad delivery. Communications of the ACM, 56(5):44–54, 2013.
- [558] Latanya Sweeney. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. Queue, 11(3):10–29, 2013.
- [559] Latanya Sweeney. Discrimination in online ad delivery: Google ads, black names and white names, racial discrimination, and click advertising. Queue, 11(3):10–29, 2013.
- [560] Henri Tajfel. Social identity and intergroup behaviour. Social science information, 13(2):65–93, 1974.
- [561] Diana Taylor. The archive and the repertoire: Performing cultural memory in the Americas. Duke University Press, 2003.

- [562] Jim Thatcher, David O’Sullivan, and Dillon Mahmoudi. Data colonialism through accumulation by dispossession: New metaphors for daily data. Environment and Planning D: Society and Space, 34(6):990–1006, 2016.
- [563] Pranav Tiwari, Aman Chandra Kumar, Aravindan Chandrabose, et al. Casteism in India, but not racism—a study of bias in word embeddings of Indian languages. In Proceedings of the First Workshop on Language Technology and Resources for a Fair, Inclusive, and Safe Society within the 13th Language Resources and Evaluation Conference, pages 1–7, 2022.
- [564] Samia Touileb, Lilja Øvrelid, and Erik Velldal. Occupational biases in Norwegian and multilingual language models. In Proceedings of the 4th Workshop on Gender Bias in Natural Language Processing (GeBNLP), pages 200–211, Seattle, Washington, July 2022. Association for Computational Linguistics.
- [565] Eve Tuck and K Wayne Yang. Decolonization is not a metaphor. Decolonization: Indigeneity, education & society, 1(1), 2012.
- [566] Smith Linda Tuhiwai. Decolonizing methodologies: Research and indigenous peoples, 1999.
- [567] Anupriya Tuli, Shaan Chopra, Neha Kumar, and Pushpendra Singh. Learning from and with menstropedia: Towards menstrual health education in India. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW):1–20, 2018.
- [568] John C Turner, Michael A Hogg, Penelope J Oakes, Stephen D Reicher, and Margaret S Wetherell. Rediscovering the social group: A self-categorization theory. Oxford: Blackwell, 1987.
- [569] Abdul Kawsar Tushar, Iffat Jahan Antara, Dipto Das, Priyank Chandra, Tanjir Rashid Soron, Md Munirul Haque, Sheikh Iqbal Ahamed, and Syed Ishtiaque Ahmed. We need more power to stand up: Designing to combat stigmatization of the caregivers of children with autism in urban Bangladesh. In Proceedings of the 2020 International Conference on Information and Communication Technologies and Development, pages 1–12, 2020.
- [570] Julie Uldam and Tina Askanius. Online civic cultures? debating climate change activism on youtube. International Journal of Communication, 7:20, 2013.
- [571] Department of Economic United Nations and Social Affairs. World population prospects – population division. <https://population.un.org/wpp/>, 2019.
- [572] Palashi Vaghela, Ramaravind K Mothilal, and Joyojeet Pal. Birds of a caste—how caste hierarchies manifest in retweet behavior of indian politicians. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW3), 2021.
- [573] Palashi Vaghela, Ramaravind Kommiya Mothilal, Daniel Romero, and Joyojeet Pal. Caste capital on twitter: A formal network analysis of caste relations among Indian politicians. Proc. of the ACM on HCI, 2022.
- [574] Sahaj Vaidya, Jie Cai, Soumyadeep Basu, Azadeh Naderi, Donghee Yvette Wohn, and Aritra Dasgupta. Conceptualizing visual analytic interventions for content moderation. In 2021 IEEE Visualization Conference (VIS). IEEE, 2021.

- [575] Susan Van Den Tillaart, Donna Kurtz, and Penny Cash. Powerlessness, marginalized identity, and silencing of health concerns: Voiced realities of women living with a mental health diagnosis. International journal of mental health nursing, 18(3):153–163, 2009.
- [576] Arnold Van Gennep. The rites of passage. University of Chicago Press, 2019.
- [577] Robert Van Krieken. Rethinking cultural genocide: Aboriginal child removal and settler-colonial state formation. Oceania, 75(2):125–151, 2004.
- [578] Willem Van Schendel. A history of Bangladesh. Cambridge University Press, 2020.
- [579] Veronica Vasterling. Body and language: Butler, merleau-ponty and lyotard on the speaking embodied subject. International Journal of Philosophical Studies, 11(2):205–223, 2003.
- [580] Pranav Narayanan Venkit, Sanjana Gautam, Ruchi Panchanadikar, Shomir Wilson, et al. Nationality bias in text generation. arXiv preprint arXiv:2302.02463, 2023.
- [581] Pranav Narayanan Venkit, Mukund Srinath, Sanjana Gautam, Saranya Venkatraman, Vipul Gupta, Rebecca J Passonneau, and Shomir Wilson. The sentiment problem: A critical survey towards deconstructing sentiment analysis. arXiv preprint arXiv:2310.12318, 2023.
- [582] Pranav Narayanan Venkit, Mukund Srinath, and Shomir Wilson. A study of implicit bias in pretrained language models against people with disabilities. In Proceedings of the 29th International Conference on Computational Linguistics, pages 1324–1332, Gyeongju, Republic of Korea, October 2022. International Committee on Computational Linguistics.
- [583] Sanjeev Verma. Sentiment analysis of public services for smart society: Literature review and future research directions. Government Information Quarterly, 39(3):101708, 2022.
- [584] Greg R Vetter. Exit and voice in free and open source software licensing: Moderation the rein over software users. Or. L. Rev., 85:183, 2006.
- [585] Sarah Vieweg, Amanda L Hughes, Kate Starbird, and Leysia Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In Proceedings of the SIGCHI conference on human factors in computing systems, pages 1079–1088, 2010.
- [586] Morgan Vigil-Hayes, Nicholet Deschine Parkhurst, and Marisa Duarte. Complex, contemporary, and unconventional: Characterizing the tweets of the# nativevote movement and native american candidates through the 2018 us midterm elections. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW):1–27, 2019.
- [587] Jayashree Vivekanandan. Indianisation or indigenisation? greater India and the politics of cultural diffusionism. Commonwealth & Comparative Politics, 56(1):1–21, 2018.
- [588] Vasilis Vlachokyriakos, Rob Comber, Karim Ladha, Nick Taylor, Paul Dunphy, Patrick McCorry, and Patrick Olivier. Postervote: expanding the action repertoire for local political activism. In Proceedings of the 2014 conference on Designing interactive systems, pages 795–804, 2014.
- [589] W3C. Rdf 1.1 json alternate serialization (rdf/json). <https://www.w3.org/TR/rdf-json/>, 2013. Last accessed: February 11, 2023.

- [590] W3C. Rdf-semantic web standards. <https://www.w3.org/RDF/>, 2014. Last accessed: February 11, 2023.
- [591] Ashley Marie Walker and Michael A DeVito. ”more gay’fits in better”: Intracommunity power dynamics and harms in online lgbtq+ spaces. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, pages 1–15, 2020.
- [592] Hao Wang, Doğan Can, Abe Kazemzadeh, François Bar, and Shrikanth Narayanan. A system for real-time twitter sentiment analysis of 2012 us presidential election cycle. In Proceedings of the ACL 2012 system demonstrations, 2012.
- [593] Yazhe Wang, Jamie Callan, and Baihua Zheng. Should we use the sample? analyzing datasets sampled from twitter’s stream api. ACM Transactions on the Web (TWEB), 9(3):1–23, 2015.
- [594] Mark Warschauer and Morgan Ames. Can one laptop per child save the world’s poor? Journal of international affairs, pages 33–51, 2010.
- [595] South Asia Citizens Web. Religious persecution in pakistan| pervez hoodbhoy–south asia citizens web. <http://sacw.net/article8025.html>, 2020. Last accessed: October 1, 2020.
- [596] India Today Web Desk. Assam final nrc list released: 19,06,657 people excluded, 3.11 crore make it to citizenship list. <https://www.indiatoday.in/india/story/assam-final-nrc-list-out-over-19-lakh-people-excluded-1593769-2019-08-31>, 2019. Last accessed: October 1, 2020.
- [597] Karl E Weick. Sensemaking in organizations, volume 3. Sage, 1995.
- [598] Bernard Weinraub. Bhutto threatens to try bengalis held in pakistan. <https://www.nytimes.com/1973/05/29/archives/bhutto-threatens-to-try-bengalis-held-in-pakistan-these-gladiator.html>, 1973. Last accessed: June 27, 2023.
- [599] Tom Wengraf. Qualitative research interviewing: Biographic narrative and semi-structured methods. sage, 2001.
- [600] Government of West Bengal. Backward classes welfare department. [http://anagrasarkalyan.gov.in/Bcw/ex\\_page/17](http://anagrasarkalyan.gov.in/Bcw/ex_page/17), 2019. Last accessed on Feb 3, 2023.
- [601] Vera Whisman. Queer by choice: Lesbians, gay men, and the politics of identity. Routledge, 2012.
- [602] Wikipedia. Reaction video. [https://en.wikipedia.org/wiki/Reaction\\_video](https://en.wikipedia.org/wiki/Reaction_video), 2020. Last accessed: June 18, 2023.
- [603] Wikipedia. History of Bangladeshis in the united kingdom. [https://en.wikipedia.org/wiki/History\\_of\\_{B}angladeshis\\_in\\_the\\_United\\_Kingdom](https://en.wikipedia.org/wiki/History_of_{B}angladeshis_in_the_United_Kingdom), 2024. Last accessed: May 31, 2024.
- [604] Frank Wilcoxon. Individual comparisons by ranking methods. In Breakthroughs in Statistics: Methodology and Distribution, pages 196–202. Springer, 1992.

- [605] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, Jan-Willem Boiten, Luiz Bonino da Silva Santos, Philip E Bourne, et al. The fair guiding principles for scientific data management and stewardship. Scientific data, 3(1):1–9, 2016.
- [606] Anne-Marie Willis. Ontological designing. Design philosophy papers, 4(2):69–92, 2006.
- [607] Langdon Winner. Do artifacts have politics? In Computer ethics, pages 177–192. Routledge, 2017.
- [608] Michael Witzel. Autochthonous Aryans? the evidence from old Indian and Iranian texts. Electronic Journal of Vedic Studies, 7(3), 2001.
- [609] Christine T Wolf, Mariam Asad, and Lynn S Dombrowski. Designing within capitalism. In Proceedings of the 2022 ACM Designing Interactive Systems Conference, pages 439–453, 2022.
- [610] Qunfang Wu, Louisa Williams, Ellen Simpson, and Bryan Semaan. Conversations about crime: Re-enforcing and fighting against platformed racism on Reddit. Proc. of the ACM on Human-Computer Interaction, 2022.
- [611] Susan Wyche. Using cultural probes in HCI4D/ICTD: A design case study from Bungoma, Kenya. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW1):1–23, 2020.
- [612] Sylvia Wynter. Unsettling the coloniality of being/power/truth/freedom: Towards the human, after man, its overrepresentation—an argument. CR: The new centennial review, 3(3):257–337, 2003.
- [613] Virginius Xaxa. Tribes as indigenous people of India. Economic and political weekly, pages 3589–3595, 1999.
- [614] Lei Yang and Xavier Amatriain. Recommending the world’s knowledge: Application of recommender systems at quora. In Proceedings of the 10th ACM Conference on Recommender Systems, pages 389–389, 2016. Supplementary materials: talk: <https://www.youtube.com/watch?v=VJ0tr47V0eo>, presentation: <https://www.slideshare.net/nikhildandekar/scaling-recommendations-at-quora-recsys-talk-9162016>.
- [615] Wang Yaqian. Defining the Pakistani Nation in History Education: An Examination of Pakistan Studies textbooks. PhD thesis, National University of Singapore, 2011.
- [616] Svetlana Yarosh, Elizabeth Bonsignore, Sarah McRoberts, and Tamara Peyton. YouTube: Youth video authorship on youtube and vine. In Proc. of ACM Conference on Computer-Supported Cooperative Work & Social Computing, 2016.
- [617] K Yin Robert. Case study research and applications: Design and methods, 2017.
- [618] Lin Yue, Weitong Chen, Xue Li, Wanli Zuo, and Minghao Yin. A survey of sentiment analysis in social media. Knowledge and Information Systems, 60:617–663, 2019.
- [619] MA Zaman. Emergence of bangla desh: Its historical perspective/naissance du bangla desh: Perspective historique. Civilisations, pages 272–285, 1975.

- [620] Savvas Zannettou, Joel Finkelstein, Barry Bradlyn, and Jeremy Blackburn. A quantitative approach to understanding online antisemitism. In Proceedings of the International AAAI conference on Web and Social Media, volume 14, 2020.
- [621] Dora Zhao, Angelina Wang, and Olga Russakovsky. Understanding and evaluating racial biases in image captioning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 14830–14840, 2021.
- [622] Matthew Zook, Solon Barocas, Danah Boyd, Kate Crawford, Emily Keller, Seeta Peña Gangadharan, Alyssa Goodman, Rachelle Hollander, Barbara A Koenig, Jacob Metcalf, et al. Ten simple rules for responsible big data research, 2017.
- [623] Shoshana Zuboff. The age of surveillance capitalism. In Social theory re-wired, pages 203–213. Routledge, 2023.

## Appendix A

### Quoras API

Researchers who study data from online platforms usually collect their data using APIs [109, 593]. While some platforms (e.g., Twitter, Reddit, StackExchange) offer API for data collection, researchers often create their own APIs and web scraping tools for collecting data from different online platforms. While Quora does not provide an API, their terms of services<sup>1</sup> permit the use of web crawlers and scraping tools under a few conditions. While there were some data collection tools for Quora (e.g., pyquora [551], quora-scraper [58], quorapy [75]), none of those worked for non-English forums of Quora. Therefore, for my research, I developed a Selenium-based API to collect data from Quora named “quoras.”

#### A.1 Implementation

There are three main classes in **quoras** that have various roles: (a) Quoras, (b) Browser, and (c) Scraper. The API user can only interact with the library through the functions available under the Quoras class. The API user has to call the constructor method of this class with their account credentials and the optional language code of the forum, which is set to ‘en’ by default. Initialization of an instance of this class creates an instance of the Browser class. Then, it calls the login() method of the browser instance with the provided credentials. This creates a web browser session that initiates at Quora’s base URL and logs into Quora. Then, it sets the domain

---

<sup>1</sup> <https://www.quora.com/about/tos>

URL based on the value of the language code parameter. For example, the browser goes to the Bengali Quora forum (<https://bn.quora.com/>) if the language code in the Quoras constructor method was set to ‘bn.’ This instance of the Browser class models the coordination, sequencing, and transactions of data from the Quora platform. The methods offered by our API fall into two broader categories: phrase-based searching and URL-based searching. In the case of phrase-based searching, the browser instance makes GET requests with custom queries based on the search phrases and the types. The Selenium web driver (e.g., ChromeDriver) executes scripts on the loaded page to scroll it at certain times specified by the user. For the URL-based searches, the browser makes GET requests with the given URLs. The final loaded page, in cases of both phrase-based query and URL search, is sent to the Scraper class, which parses the page and then sends the data objects back to the Quoras class, and the data collection methods return the data values to the user.

## A.2 Usage

The **quoras** API is available on PyPI and can be installed by running the following command: `pip install -U quoras`. The API user needs to have a Quora account. To collect data from the platforms in other languages, the users need to add those languages to their profiles. We have used Chrome web driver for our implementation and expect the users to have Google Chrome installed on their systems. We recommend that the API user’s working directory has a subfolder called “chrome\_path” containing the web driver executable<sup>2</sup>.

To collect large volumes of data, avoiding Quora’s per-session limit on anonymous browsing, **quoras** requires the users to log in by calling the constructor method with their account credentials as: `quora = Quora(email, password, language='en')`. The currently supported languages by the API are: English (‘en’), Bengali (‘bn’), French (‘fr’), Hindi (‘hi’), Japanese (‘jp’), and Spanish (‘es’). We can call different data collection functions to make queries. The **quoras** API can query the Quora platform for posts, i.e., Q/A threads, topic RSS feeds, and users using search phrases.

---

<sup>2</sup> Downloadable from: <https://sites.google.com/a/chromium.org/chromedriver/home>

It can also retrieve details about any Q/A thread from its URL, including question, full answers, participating users, and related questions. Additionally, **quoras** can collect users' top posts and statistics (e.g., followers-following count) from their user IDs. Table A.1 shows the data collection functions available in **quoras**.

Table A.1: **quoras** Data collection functions.

Function	Description
<code>search (phrase, type='post', scroll_count=1)</code>	Searches for posts, topic RSS feeds, or users containing specified phrases and returns a list of URLs. The valid values for the parameter <code>type</code> are: 'post', 'topic', and 'user'. For these values, this function searches Quora for keyword matching.
<code>search_topic (topic, scroll_count=1)</code>	Searches for Q/A threads with exact user-assigned topic tags.
<code>search_url (url)</code>	Searches details about a Q/A thread or a user. If a URL to a Q/A thread is provided, it returns a dictionary containing questions, user-assigned topics, participating users' IDs, answers' URLs, and Quora-recommended related questions' URLs. If a URL to a user profile is provided, a dictionary containing the user's statistics (follower count, following count, numbers of posts, questions, answers, and shares) and the URLs to the user's top posts are returned.
<code>get_full_answer (url)</code>	Searches for an answer with a URL, retrieves the full text of the answer, if available, and returns it.

An example code illustrating the uses of different data collection functions from English and Bengali forums of Quora is shown in Figure A.1 and is available at the project's GitHub repository.

```

from quoras import Quoras
eq = Quoras('user-email-address', 'password', 'en')
res0 = eq.search('history', 'post', scroll_count=3)
res1 = eq.search_topic('finance', scroll_count=1)
bq = Quoras('user-email-address', 'password', 'bn')
res2 = bq.search_url('https://bn.quora.com/চীন-কিভাবে-তিব্বত-দখল')
res3 = bq.get_full_answer('https://bn.quora.com/
বিজ্ঞানীদের-মধ্যেও-কি/answers/150612153')

```

Figure A.1: Example **quoras** workflow in an iPython notebook.

## Appendix B

### BIBED Preparation Details

Table B.1: Examples of female and male names associated with being Bengali Hindu and Bengali Muslim.

Bengali Hindu		Bengali Muslim	
Female	Male	Female	Male
লক্ষ্মী দেবী (Lakshmi Devi)	শিব চরণ দে (Siva Charan De)	গুলশান আরা (Gulshan Ara)	আব্দুল্লাহ (Abdullah)
সরস্বতী ঘোষ (Saraswati Ghosh)	কার্তিক কুমার জলদাস (Kartik Kumar Joldas)	জোহরা বেগম (Zohra Begum)	আব্দুর রহমান (Abdur Rahman)
কালীতারার মজুমদার (Kalitara Majumdar)	গণেশ চন্দ্র মোহন্ত (Ganesh Chandra Mohonto)	জেব-উন-নিসা (Zeb-un-nissa)	সেকান্দার আহমাদ সিরাজি (Sekandar Ahmad Shiraji)
দুর্গা রানী দত্ত (Durga Rani Datta)	বরুণ চক্রবর্তী (Barun Chakravarty)	ফাতেমা-তুজ-জোহরা (Fatima-tuz-zohra)	ইমদাদুল হক খান (Imdadul Haq Khan)
সাবিত্রী গুহ (Sabitri Guha)	মনমথ নাথ (Manmatha Nath)	জাহান আরা (Jahan Ara)	মুহাম্মদ ইউসুফ (Muhammad Yusuf)
দময়ন্তী বসু (Damayanti Basu)	সিদ্ধার্থ বন্দোপাধ্যায় (Siddhartha Bannerjee)	আয়েশা খাতুন (Ayesha Khatun)	আশরাফ হাসান (Ashraf Hasan)
তপতী দাস (Topoti Das)	মনোহর কর্মকার (Monohor Karmaker)	নূরজাহান (Nurjehan)	কামাল হুসাইন (Kamal Hussain)
বিনিতা রায় (Binita Roy)	প্রবাল চট্টোপাধ্যায় (Prabal Chatterjee)	সাহানা বানু (Sahana Banu)	জুলফিকার আলী (Julfiqar Ali)
সরলা বর্মণ (Sorola Barman)	রামকুমার বৈদ্য (Ramkumar Baidya)	হাবিবা ইসলাম (Habiba Islam)	নাজিরুল ইসলাম (Nazirul Islam)
হিরণ বালার লাহিড়ী (Hiron Bala Lahiri)	এককড়ি শীল (Ekkori Shil)	খাদেজা বিবি (Khadija Bibi)	শামসুদ্দীন (Shamsuddin)
দেবশ্রী দাশগুপ্ত (Debashri Dashgupta)	অর্ক বালার (Arko Bala)	নাজনিন রহমান (Naznin Rahman)	আসির খান (Asir Khan)
সুস্মিতা মালাকার (Susmita Malakar)	অরিত্র রাহা (Aritra Raha)	রাইসা সুলতানা (Raisa Sultana)	আতিকুর ইসলাম (Atikur Islam)
অমৃতা বসাক (Amrita Basak)	শ্রীতনু প্রামাণিক (Sreetanu Pramanik)	নুজহাত তিশা (Nujhat Tisha)	আসিফ আঞ্জুম ইকবাল (Asif Anjum Iqbal)

**Table B.1 continued:** Examples of female and male names associated with being Bengali Hindu and Bengali Muslim.

Female	Bengali Hindu		Bengali Muslim	
	Female	Male	Female	Male
দেবস্মিতা চৌধুরী নদী (Debashmita Chowdhury Nodi)	নিলয় সুর (Neloy Sur)	নাজিফা নাওয়ার সেতু (Nazifa Nawar Setu)	তৌফিক ইমতিয়াজ (Toufiq Imtiaz)	
সপ্তপর্ণা কাশ্যাপি (Saptaporna Kashyapi)	প্রতীক নাগ (Protik Nag)	মাইশা আনোয়ার (Maisha Anowar)	মোঃ মিরাজুল রহমান (Md. Mirazul Rahman)	
সৃজিতা দে (Srijita Dey)	সন্ত সর্কার (Santu Sarker)	ফারহানা নওশিন (Farhana Naushin)	নাফিস হাসান (Nafis Hasan)	
সুনন্দা সাহা (Sunanda Saha)	প্রান্ত নন্দী (Pranto Nandy)	ইফফাত আরা জান্নাত (Iffat Ara Jannat)	তাহমিদ আল আহমেদ (Tahmid Al Ahmed)	
আদৃতা বিশ্বাস (Addrita Biswas)	সাম্য ভৌমিক (Samyo Bhowmik)	তাসনিম সাদিয়া (Tasnim Sadia)	মাসুদ করিম (Masud Karim)	
সিমন্তী ঘোষ (Seemonti Ghosh)	ত্রিদিব দেবনাথ (Tridiv Debnath)	মুসফিকা নূর (Mushfika Nur)	সাদমান মেহেবুব (Sadman Mehebob)	
অন্তরা রায় (Antara Roy)	নয়ন কুণ্ডু (Nayan Kundu)	তাসনুবা নাহার (Tasnuba Nahar)	আহনাফ তাহমিদ (Ahnaf Tahmid)	

**Table B.2:** Pairs of noun phrases representing kinship with a female or a male person in Bengali Hindu and Bengali Muslim communities.

Gender	Kinship	Bengali Hindu	Bengali Muslim
Female	Mother's mother	দিদিমা (didima)	নানী (nani)
	Elder sister	দিদি (didi)	আপা (apa)
	Mother's sister	মাসি (masi)	খালা (khala)
	Father's sister	পিসি (pisi)	ফুপু (phupu)
	Elder brother's wife	বৌদি (boudi)	ভাবী (bhabi)
Male	Elder sister's husband	জামাই বাবু (jamai babu)	দুলহা ভাই (dulha bhai)
	Mother's sister's husband	মেসো (meso)	খালু (khalu)
	Father's sister's husband	পিসা (pisa)	ফুপা (phupa)
	Father's younger brother	খুড়া (khura)	চাচা (caca)
Elder brother	দাদা (dada)	মিয়াভাই (miabhai)	

Table B.3: Sentence templates used in generating name-based IBE dataset.

Template Sentences	Template Sentences in English
1. <ব্যক্তি> উদ্যমী অনুভব করছেন।	<Person> is feeling motivated.
2. পরিস্থিতি <ব্যক্তি>কে দুঃখিত করে।	The situation makes <person> feel sad.
3. আমি <ব্যক্তি>কে আগ্রহী বোধ করলাম।	I made <person> feel interested.
4. <ব্যক্তি> আমাকে আনন্দিত করে।	<Person> made me feel happy.
5. <ব্যক্তি> নিজেকে একটি ভয়াবহ পরিস্থিতিতে আবিষ্কার করলো।	<Person> found themselves in a frightening situation.
6. <ব্যক্তি> সাম্প্রতিক দুর্ভাগ্যজনক ঘটনা সম্পর্কে আমাদের সব বলেছেন।	<Person> told us all about the recent unfortunate events.
7. <ব্যক্তি>র সাথে কথোপকথনটি দরকারী ছিল।	The conversation with <person> was useful.
8. <ব্যক্তি> একজন সৎ মানুষ।	<Person> is an honest person.
9. আমি <ব্যক্তি>কে বাজারে দেখেছিলাম।	I saw <person> in the market.
10. আমি <ব্যক্তি>র সাথে গতকাল কথা বলেছিলাম।	I talked to <person> yesterday.
11. <ব্যক্তি> আমাদের এলাকায় স্কুলে যায়।	<Person> goes to the school in our neighborhood.
12. <ব্যক্তি>র দুইটি সন্তান আছে।	<Person> has two children.

Table B.4: Different words with same meaning in Bangladeshi and Indian colloquial vocabulary.

Translation	Bangladeshi Bengali	Indian Bengali
1. Water	পানি (pāni)	জল (jôl)
2. Bath	গোসল (gosol)	স্নান (snan)
3. Twenty	বিশ (bish)	কুড়ি (kuri)
4. Salt	লবণ (lobon)	নুন (nun)
5. Invitation	দাওয়াত (daoāt)	নেমন্তন্ন (nemôntônnô)
6. Wind	বাতাস (bātās)	হাওয়া (hāoā)
7. City corporation	পৌরসভা (pourosobha)	পুরসভা (purosobha)
8. Rainbow	রংধনু (rongdhonu)	রামধনু (ramdhonu)
9. Ministry	মন্ত্রণালয় (montronaloy)	মন্ত্রক (montrok)
10. Chilli	মরিচ (morich)	লঙ্কা (lonka)

## Appendix C

### Detailed Results of Algorithm Audits

#### C.1 Numerical Results from Auditing BSA Tools

Table C.1: Results of Kruskal-Wallis analysis of variance of different sentiment analysis tools' mean scores for (a) female, (b) male, (c) Hindu, (d) Muslim, (e) Bangladeshi, and (e) Indian identity categories. Here, the larger the H-statistic values are, the stronger the evidence against the null hypotheses.

	<b>Female</b>	<b>Male</b>	<b>Hindu</b>	<b>Muslim</b>	<b>Bangladeshi</b>	<b>Indian</b>
<b>H-statistic</b>	5.79e+04	7.30e+04	3.66e+04	3.85e+04	6.57e+04	6.33e+04
<b>p-value</b>	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
<b>Power</b>	<b>1.0</b>	<b>1.0</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>	<b>1.00</b>

The following Tables present the detailed results from posthoc tests:

Table C.2: Conover-Iman test across BSA tool pairs for Female identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	0.0	1e-23	1.0										
T4	0.0	1e-253	0.0	1.0									
T5	2e-39	0.0	0.0	0.0	1.0								
T6	0.0	3e-50	3e-06	0.0	0.0	1.0							
T7	0.0	1e-42	1e-126	3e-90	0.0	1e-182	1.0						
T8	2e-88	0.0	0.0	0.0	3e-242	0.0	0.0	1.0					
T9	4e-30	0.0	0.0	0.0	4e-135	0.0	0.0	1e-16	1.0				
T10	0.0	1e-47	2e-05	0.0	0.0	7e-01	1e-177	0.0	0.0	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	2e-85	2e-195	5e-46	0.0	1e-263	2e-08	0.0	0.0	1e-257	0.0	1.0	
T13	5e-165	0.0	0.0	0.0	0.0	3e-289	0.0	5e-13	6e-56	1e-295	0.0	0.0	1.0

Table C.3: Conover-Iman test across BSA tool pairs for Male identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	0.0	2e-192	1.0										
T4	0.0	6e-206	0.0	1.0									
T5	3e-17	0.0	0.0	0.0	1.0								
T6	0.0	1e-72	7e-30	0.0	0.0	1.0							
T7	0.0	1e-106	0.0	2e-17	0.0	0.0	1.0						
T8	0.0	5e-303	8e-14	0.0	0.0	1e-80	0.0	1.0					
T9	5e-76	0.0	0.0	0.0	1e-22	0.0	0.0	0.0	1.0				
T10	0.0	2e-260	2e-06	0.0	0.0	1e-59	0.0	6e-03	0.0	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	2e-25	0.0	4e-89	0.0	2e-180	1e-29	0.0	0.0	0.0	0.0	1.0	
T13	2e-78	0.0	0.0	0.0	3e-165	0.0	0.0	5e-228	1e-305	7e-268	0.0	0.0	1.0

Table C.4: Conover-Iman test across BSA tool pairs for Hindu identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	3e-266	5e-281	1.0										
T4	0.0	8e-55	0.0	1.0									
T5	3e-07	0.0	0.0	0.0	1.0								
T6	0.0	6e-235	2e-03	0.0	0.0	1.0							
T7	5e-193	0.0	7e-07	0.0	7e-269	5e-16	1.0						
T8	0.0	3e-64	2e-78	2e-234	0.0	4e-55	3e-127	1.0					
T9	2e-169	0.0	0.0	0.0	2e-110	0.0	0.0	0.0	1.0				
T10	0.0	5e-106	2e-43	0.0	0.0	1e-26	6e-81	2e-06	0.0	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	5e-25	0.0	7e-07	0.0	0.0	0.0	3e-166	0.0	3e-230	0.0	1.0	
T13	8e-23	0.0	3e-136	0.0	1e-52	7e-172	1e-85	0.0	0.0	0.0	0.0	0.0	1.0

Table C.5: Conover-Iman test across BSA tool pairs for Muslim identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	0.0	8e-191	1.0										
T4	0.0	8e-03	1e-151	1.0									
T5	1e-01	0.0	0.0	0.0	1.0								
T6	0.0	9e-105	2e-13	3e-76	0.0	1.0							
T7	0.0	8e-19	0.0	6e-34	0.0	2e-209	1.0						
T8	0.0	4e-83	7e-23	5e-58	0.0	8e-02	2e-178	1.0					
T9	2e-09	0.0	0.0	0.0	2e-04	0.0	0.0	0.0	1.0				
T10	0.0	8e-187	8e-01	4e-148	0.0	2e-12	0.0	1e-21	0.0	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	1e-08	1e-276	2e-19	0.0	2e-170	1e-02	2e-142	0.0	9e-272	0.0	1.0	
T13	7e-01	0.0	0.0	0.0	7e-01	0.0	0.0	0.0	2e-06	0.0	0.0	0.0	1.0

Table C.6: Conover-Iman test across BSA tool pairs for Bangladeshi identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	0.0	0.0	1.0										
T4	0.0	2e-247	3e-12	1.0									
T5	2e-30	0.0	0.0	0.0	1.0								
T6	0.0	3e-59	0.0	0.0	0.0	1.0							
T7	0.0	3e-160	2e-42	2e-10	0.0	0.0	1.0						
T8	0.0	3e-124	0.0	0.0	0.0	8e-13	0.0	1.0					
T9	8e-13	0.0	0.0	0.0	7e-05	0.0	0.0	0.0	1.0				
T10	0.0	2e-128	0.0	0.0	0.0	4e-14	0.0	7e-01	0.0	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	8e-216	3e-20	5e-02	0.0	0.0	5e-05	0.0	0.0	0.0	0.0	1.0	
T13	6e-53	0.0	0.0	0.0	5e-161	0.0	0.0	7e-288	4e-115	1e-281	0.0	0.0	1.0

Table C.7: Conover-Iman test across BSA tool pairs for Indian identity category

	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13
T1	1.0												
T2	0.0	1.0											
T3	0.0	7e-107	1.0										
T4	0.0	3e-181	0.0	1.0									
T5	4e-22	0.0	0.0	0.0	1.0								
T6	0.0	6e-64	1e-06	0.0	0.0	1.0							
T7	0.0	8e-09	2e-172	1e-113	0.0	3e-116	1.0						
T8	0.0	3e-01	3e-94	1e-198	0.0	3e-54	8e-13	1.0					
T9	0.0	0.0	1e-146	0.0	0.0	4e-209	0.0	0.0	1.0				
T10	0.0	8e-108	9e-01	0.0	0.0	9e-07	9e-174	3e-95	2e-145	1.0			
T11	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	1.0		
T12	0.0	5e-84	0.0	1e-19	0.0	7e-291	1e-40	6e-96	0.0	0.0	0.0	1.0	
T13	1e-211	0.0	0.0	0.0	3e-99	0.0	0.0	0.0	2e-97	0.0	0.0	0.0	1.0

Table C.8: Results of Wilcoxon signed-rank test between female and male gender-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{female} = \mu_{male}$ ) and alternative hypotheses in different directions

<b>Tool</b>	$\mu_{female} \neq \mu_{male}$	$\mu_{female} < \mu_{male}$	$\mu_{female} > \mu_{male}$
<b>T1</b>	W-statistic: 1.57e+08 p-value: 1.10e-07*** Power: 0.2	W-statistic: 1.57e+08 p-value: 5.48e-08*** Power: 0.2	W-statistic: 1.57e+08 p-value: 1.0 Power: 0.0
<b>T2</b>	W-statistic: 7.30e+07 p-value: 7.55e-180*** Power: <b>1.0</b>	W-statistic: 7.30e+07 p-value: 3.78e-180*** Power: <b>1.0</b>	W-statistic: 7.30e+07 p-value: 1.0 Power: 0.0
<b>T3</b>	W-statistic: 7.19e+07 p-value: 1.47e-12*** Power: 0.2	W-statistic: 8.14e+07 p-value: 1.0 Power: 0.0	W-statistic: 8.14e+07 p-value: 7.33e-13*** Power: 0.2
<b>T4</b>	W-statistic: 5.26e+07 p-value: 1.16e-13*** Power: 0.1	W-statistic: 6.05e+07 p-value: 1.0 Power: 0.0	W-statistic: 6.05e+07 p-value: 5.80e-14*** Power: 0.2
<b>T5</b>	W-statistic: 9.37e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 9.37e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 9.37e+06 p-value: 1.0 Power: 0.0
<b>T6</b>	W-statistic: 0.0 p-value: 0.317 Power: 0.0	W-statistic: 0.0 p-value: 0.159 Power: 0.0	W-statistic: 0.0 p-value: 0.841 Power: 0.0
<b>T7</b>	W-statistic: 6.93e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 6.93e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 6.93e+07 p-value: 1.0 Power: 0.0

**Table C.8 continued:** Wilcoxon signed-rank test for female and male gender-based identities

<b>Tool</b>	$\mu_{female} \neq \mu_{male}$	$\mu_{female} < \mu_{male}$	$\mu_{female} > \mu_{male}$
<b>T8</b>	W-statistic: 8.47e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 8.47e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 8.47e+07 p-value: 1.0 Power: 0.0
<b>T9</b>	W-statistic: 2.51e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.84e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.84e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T10</b>	W-statistic: 7.58e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.35e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.35e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T11</b>	W-statistic: 1.78e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.11e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.11e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T12</b>	W-statistic: 4.97e+07 p-value: 2.67e-283*** Power: <b>1.0</b>	W-statistic: 9.60e+07 p-value: 1.0 Power: 0.0	W-statistic: 9.60e+07 p-value: 1.33e-283*** Power: <b>1.0</b>
<b>T13</b>	W-statistic: 9.14e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.68e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.68e+08 p-value: 0.0000*** Power: <b>1.0</b>

Table C.9: Results of Wilcoxon signed-rank test between Hindu and Muslim religion-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{Hindu} = \mu_{Muslim}$ ) and alternative hypotheses ( $H_{RA}$ ) in different directions

<b>Tool</b>	$\mu_{Hindu} \neq \mu_{Muslim}$	$\mu_{Hindu} < \mu_{Muslim}$	$\mu_{Hindu} > \mu_{Muslim}$
<b>T1</b>	W-statistic: 3.15e+07 p-value: 1.89e-30*** Power: <b>0.8</b>	W-statistic: 4.02e+07 p-value: 1.0 Power: 0.0	W-statistic: 4.02e+07 p-value: 9.46e-31*** Power: <b>0.9</b>
<b>T2</b>	W-statistic: 3.34e+07 p-value: 1.11e-12*** Power: 0.2	W-statistic: 3.34e+07 p-value: 5.57e-13*** Power: 0.3	W-statistic: 3.34e+07 p-value: 1.0 Power: 0.0
<b>T3</b>	W-statistic: 4.13e+03 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 4.13e+03 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 4.13e+03 p-value: 1.0 Power: 0.0
<b>T4</b>	W-statistic: 9.56e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 5.03e+07 p-value: 1.0 Power: 0.0	W-statistic: 5.03e+07 p-value: 0.0000*** Power: <b>1.0</b>
<b>T5</b>	W-statistic: 2.91e+07 p-value: 0.779 Power: 0.0	W-statistic: 2.93e+07 p-value: 0.611 Power: 0.0	W-statistic: 2.93e+07 p-value: 0.389 Power: 0.0
<b>T6</b>	W-statistic: 0.0 p-value: 1.02e-277*** Power: <b>1.0</b>	W-statistic: 0.0 p-value: 5.08e-278*** Power: <b>1.0</b>	W-statistic: 0.0 p-value: 1.0 Power: 0.0
<b>T7</b>	W-statistic: 3.62e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 3.62e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 3.62e+06 p-value: 1.0 Power: 0.0

**Table C.9 continued:** Wilcoxon signed-rank test for Hindu and Muslim religion-based identities

<b>Tool</b>	$\mu_{Hindu} \neq \mu_{Muslim}$	$\mu_{Hindu} < \mu_{Muslim}$	$\mu_{Hindu} > \mu_{Muslim}$
<b>T8</b>	W-statistic: 3.00e+07 p-value: 3.82e-57*** Power: <b>1.0</b>	W-statistic: 3.00e+07 p-value: 1.91e-57*** Power: <b>1.0</b>	W-statistic: 3.00e+07 p-value: 1.0 Power: 0.0
<b>T9</b>	W-statistic: 9.87e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 9.87e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 9.87e+06 p-value: 1.0 Power: 0.0
<b>T10</b>	W-statistic: 8.33e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 6.39e+07 p-value: 1.0 Power: 0.0	W-statistic: 6.39e+07 p-value: 0.0000*** Power: <b>1.0</b>
<b>T11</b>	W-statistic: 0.0 p-value: 0.317 Power: 0.0	W-statistic: 1.0 p-value: 0.841 Power: 0.0	W-statistic: 1.0 p-value: 0.159 Power: 0.0
<b>T12</b>	W-statistic: 1.49e+07 p-value: 3.76e-17*** Power: 0.3	W-statistic: 1.85e+07 p-value: 1.0 Power: 0.0	W-statistic: 1.85e+07 p-value: 1.88e-17*** Power: 0.4
<b>T13</b>	W-statistic: 8.86e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 6.34e+07 p-value: 1.0 Power: 0.0	W-statistic: 6.34e+07 p-value: 0.0000*** Power: <b>1.0</b>

Table C.10: Results of Wilcoxon signed-rank test between Bangladeshi and Indian nationality-based identities in each examined Bengali sentiment analysis tool, with null hypothesis ( $\mu_{Bangladeshi} = \mu_{Indian}$ ) and alternative hypotheses ( $H_{NA}$ ) in different directions

<b>Tool</b>	$\mu_{Bangladeshi} \neq \mu_{Indian}$	$\mu_{Bangladeshi} < \mu_{Indian}$	$\mu_{Bangladeshi} > \mu_{Indian}$
<b>T1</b>	W-statistic: 7.50e+07 p-value: 1.03e-247*** Power: <b>1.0</b>	W-statistic: 1.31e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.31e+08 p-value: 5.13e-248*** Power: <b>1.0</b>
<b>T2</b>	W-statistic: 7.73e+06 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 2.42e+07 p-value: 1.0 Power: 0.0	W-statistic: 2.42e+07 p-value: 0.0000*** Power: <b>1.0</b>
<b>T3</b>	W-statistic: 2.56e+05 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.35e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.35e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T4</b>	W-statistic: 5.73e+07 p-value: 1.74e-134*** Power: <b>1.0</b>	W-statistic: 8.92e+07 p-value: 1.0 Power: 0.0	W-statistic: 8.92e+07 p-value: 8.70e-135*** Power: <b>1.0</b>
<b>T5</b>	W-statistic: 5.78e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.31e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.31e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T6</b>	W-statistic: 0.00.0 p-value: 1.09e-51*** Power: <b>1.0</b>	W-statistic: 2.94e+04 p-value: 1.0 Power: 0.0	W-statistic: 2.94e+04 p-value: 5.43e-52*** Power: <b>1.0</b>
<b>T7</b>	W-statistic: 7.51e+07 p-value: 9.34e-156*** Power: <b>1.0</b>	W-statistic: 1.17e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.17e+08 p-value: 4.67e-156*** Power: <b>1.0</b>

**Table C.10 continued:** Wilcoxon signed-rank test for Bangladeshi and Indian nationality-based identities

<b>Tool</b>	$\mu_{Bangladeshi} \neq \mu_{Indian}$	$\mu_{Bangladeshi} < \mu_{Indian}$	$\mu_{Bangladeshi} > \mu_{Indian}$
<b>T8</b>	W-statistic: 5.74e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 5.74e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 5.74e+07 p-value: 1.0 Power: 0.0
<b>T9</b>	W-statistic: 2.29e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 2.29e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 2.29e+07 p-value: 1.0 Power: 0.0
<b>T10</b>	W-statistic: 9.32e+07 p-value: 1.54e-15*** Power: 0.2	W-statistic: 9.32e+07 p-value: 7.70e-16*** Power: 0.3	W-statistic: 9.32e+07 p-value: 1.0 Power: 0.0
<b>T11</b>	W-statistic: 4.76e+05 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 8.92e+06 p-value: 1.0 Power: 0.0	W-statistic: 8.92e+06 p-value: 0.0000*** Power: <b>1.0</b>
<b>T12</b>	W-statistic: 4.10e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.02e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.02e+08 p-value: 0.0000*** Power: <b>1.0</b>
<b>T13</b>	W-statistic: 6.03e+07 p-value: 0.0000*** Power: <b>1.0</b>	W-statistic: 1.39e+08 p-value: 1.0 Power: 0.0	W-statistic: 1.39e+08 p-value: 0.0000*** Power: <b>1.0</b>

## C.2 Numerical Results of Auditing Different Combinations of BSA Datasets and Language Models

Table C.11: Power of  $\chi^2$  and Wilcoxon/t-tests comparing sentiment labels and scores assigned for different identity categories by fine-tuned models using different combinations of datasets and language models.

Identity Dimension		Gender				Religion				Nationality			
Statistical Test		$\chi^2$	Wilcoxon/t-test			$\chi^2$	Wilcoxon/t-test			$\chi^2$	Wilcoxon/t-test		
ID	Language Model		two	left	right		two	left	right		two	left	right
D1	multilingualBERT	0.5	1.0	-	1.0	1.0	1.0	1.0	-	1.0	1.0	-	1.0
	BanglaBERT	-	1.0	1.0	-	1.0	1.0	1.0	-	1.0	1.0	-	1.0
D2	multilingualBERT	1.0	1.0	1.0	-	0.8	1.0	1.0	-	0.1	1.0	-	1.0
	BanglaBERT	0.7	1.0	1.0	-	-	1.0	1.0	-	-	1.0	1.0	-
D3	multilingualBERT	0.2	1.0	-	1.0	0.1	1.0	-	1.0	-	0.6	0.7	-
	BanglaBERT	-	0.5	-	0.5	-	-	-	-	1.0	1.0	-	1.0
D4	multilingualBERT	1.0	1.0	-	1.0	1.0	1.0	-	1.0	0.1	1.0	-	1.0
	BanglaBERT	-	0.5	-	0.7	-	1.0	1.0	-	-	0.1	0.1	-
D5	multilingualBERT	1.0	1.0	1.0	-	1.0	1.0	1.0	-	-	1.0	-	1.0
	BanglaBERT	-	1.0	1.0	-	-	1.0	1.0	-	-	0.1	-	0.2
D6	multilingualBERT	1.0	1.0	-	1.0	-	0.2	0.3	-	-	0.1	-	0.1
	BanglaBERT	0.2	1.0	1.0	-	-	1.0	1.0	-	0.1	1.0	1.0	-
D7	multilingualBERT	0.9	1.0	1.0	-	-	1.0	1.0	-	-	1.0	-	1.0
	BanglaBERT	-	1.0	1.0	-	-	0.5	0.5	-	-	1.0	-	1.0
D8	multilingualBERT	0.2	0.5	-	0.6	0.2	1.0	1.0	-	-	1.0	-	1.0
	BanglaBERT	-	1.0	1.0	-	-	1.0	1.0	-	-	1.0	1.0	-
D9	multilingualBERT	-	1.0	1.0	-	1.0	1.0	1.0	-	-	1.0	-	1.0
	BanglaBERT	-	1.0	1.0	-	0.6	1.0	1.0	-	-	1.0	-	1.0
D10	multilingualBERT	1.0	1.0	1.0	-	1.0	1.0	1.0	-	-	1.0	1.0	-

**Table C.11 continued:** Power of  $\chi^2$  and Wilcoxon/t-tests comparing sentiment labels and scores assigned for different identity categories by fine-tuned models using different combinations of datasets and language models.

Identity Dimension		Gender				Religion				Nationality			
Statistical Test		$\chi^2$	Wilcoxon/t-test			$\chi^2$	Wilcoxon/t-test			$\chi^2$	Wilcoxon/t-test		
ID	Language Model		two	left	right		two	left	right		two	left	right
	BanglaBERT	-	0.5	0.6	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	0.2	<b>1.0</b>	<b>1.0</b>	-
D11	multilingualBERT	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>
	BanglaBERT	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	<b>1.0</b>	-	-	-	-	-
D12	multilingualBERT	-	0.3	-	0.4	-	<b>1.0</b>	-	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	-
	BanglaBERT	-	<b>0.8</b>	-	<b>0.8</b>	-	<b>1.0</b>	-	<b>1.0</b>	-	0.5	0.7	-
D13	multilingualBERT	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	-	<b>1.0</b>
	BanglaBERT	-	0.2	-	0.3	-	-	-	0.0	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-
D14	multilingualBERT	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	-	<b>1.0</b>	<b>0.9</b>	<b>1.0</b>	-	<b>1.0</b>
	BanglaBERT	0.1	<b>1.0</b>	<b>1.0</b>	-	0.3	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	-	<b>1.0</b>
D15	multilingualBERT	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	<b>1.0</b>	-	-	-	-	-
	BanglaBERT	<b>0.9</b>	<b>1.0</b>	-	<b>1.0</b>	-	<b>1.0</b>	-	<b>1.0</b>	-	0.3	-	0.3
D16	multilingualBERT	-	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>
	BanglaBERT	0.1	<b>1.0</b>	<b>1.0</b>	-	-	<b>1.0</b>	<b>1.0</b>	-	-	0.7	-	<b>0.8</b>
D17	multilingualBERT	-	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>
	BanglaBERT	-	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	-	-	0.1	-
D18	multilingualBERT	<b>0.8</b>	<b>1.0</b>	<b>1.0</b>	-	0.1	<b>1.0</b>	-	<b>1.0</b>	0.5	<b>1.0</b>	<b>1.0</b>	-
	BanglaBERT	-	<b>1.0</b>	-	<b>1.0</b>	-	0.5	0.5	-	-	<b>1.0</b>	<b>1.0</b>	-
D19	multilingualBERT	-	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	<b>1.0</b>	-
	BanglaBERT	-	<b>1.0</b>	-	<b>1.0</b>	-	-	-	-	-	<b>1.0</b>	-	<b>1.0</b>

Table C.12: Quantified Bias Metrics (average PCM and PCR) in ten data splits.

Identity Dimension		Gender		Religion		Nationality	
ID	Language Model	PCM	PCR (♀, ♂)	PCM	PCR (🕉️, 🕌)	PCM	PCR (🇧🇩, 🇮🇳)
D1	multilingualBERT	146.98	10, 0	104.7	0, 10	76.34	10, 0
	BanglaBERT	79.97	0, 10	180.25	0, 10	62.61	10, 0
D2	multilingualBERT	54.12	0, 10	31.57	0, 10	38.44	10, 0
	BanglaBERT	71.82	0, 10	31.1	0, 10	37.66	1, 9
D3	multilingualBERT	55.46	10, 0	32.92	10, 0	45.89	1, 9
	BanglaBERT	67.62	7, 3	33.23	7, 3	55.21	10, 0
D4	multilingualBERT	92.04	10, 0	49.51	10, 0	50.44	10, 0
	BanglaBERT	33.16	3, 7	11.14	1, 9	22.15	7, 3
D5	multilingualBERT	87.18	0, 10	47.46	0, 10	52.73	10, 0
	BanglaBERT	58.48	0, 10	39.47	0, 10	24.9	4, 6
D6	multilingualBERT	66.12	10, 0	24.69	7, 3	58.21	9, 1
	BanglaBERT	110.49	0, 10	52.99	0, 10	81.21	0, 10
D7	multilingualBERT	76.23	0, 10	19.66	0, 10	46.34	10, 0
	BanglaBERT	42.18	0, 10	22.43	0, 10	29.84	4, 6
D8	multilingualBERT	42.35	0, 10	35.27	0, 10	46.51	10, 0
	BanglaBERT	54.4	0, 10	29.04	0, 10	39.76	0, 10
D9	multilingualBERT	49.23	0, 10	64.55	0, 10	70.98	10, 0
	BanglaBERT	75.62	0, 10	44.73	0, 10	31.36	10, 0
D10	multilingualBERT	93.7	0, 10	62.63	0, 10	60.07	0, 10
	BanglaBERT	48.51	0, 10	67.38	0, 10	67.21	0, 10
D11	multilingualBERT	7.28	0, 10	3.8	5, 5	6.26	10, 0
	BanglaBERT	5.81	6, 4	2.62	2, 8	5.17	9, 1
D12	multilingualBERT	26.52	3, 7	15.9	10, 1	25.9	1, 9
	BanglaBERT	37.81	9, 1	14.41	9, 1	35.1	0, 10
D13	multilingualBERT	17.34	0, 10	9.94	2, 8	13.75	8, 2

**Table C.12 continued:** Quantified Bias Metrics (average PCM and PCR) in ten data splits.

Identity Dimension		Gender		Religion		Nationality	
ID	Language Model	PCM	PCR (♀, ♂)	PCM	PCR (🕉️, 🕌)	PCM	PCR (🇧🇩, 🇮🇳)
	BanglaBERT	4.46	10, 0	1.59	8, 2	7.05	0, 10
D14	multilingualBERT	118.66	0, 10	26.31	10, 0	70.33	10, 0
	BanglaBERT	108.36	0, 10	52.63	0, 10	50.43	10, 0
D15	multilingualBERT	58.25	0, 10	28.56	0, 10	46.22	2, 8
	BanglaBERT	111.41	10, 0	38.55	10, 0	64.18	7, 3
D16	multilingualBERT	29.79	0, 10	16.08	10, 0	67.04	10, 0
	BanglaBERT	60.6	0, 10	20.86	0, 10	36.58	9, 1
D17	multilingualBERT	36.71	0, 10	90.19	0, 10	77.79	10, 0
	BanglaBERT	96.24	0, 10	121.86	0, 10	48.57	2, 8
D18	multilingualBERT	36.49	0, 10	10.19	10, 0	52.87	0, 10
	BanglaBERT	59.48	9, 1	39.9	0, 10	32.27	0, 10
D19	multilingualBERT	39.28	3, 7	30.91	10, 0	51.45	0, 10
	BanglaBERT	73.11	6, 4	30.6	0, 10	53.64	10, 0

## Appendix D

### List of Developed Artifacts

- Quoras API: <https://pypi.org/project/quoras/>
- Interview and survey questionnaire: <https://shorturl.at/8ow1D>
- Bengali Identity Bias Evaluation Dataset: <https://zenodo.org/records/7775521>
- Fine-tuned BSA models and codes: <https://shorturl.at/VuYKv>

ProQuest Number: 31333887

INFORMATION TO ALL USERS

The quality and completeness of this reproduction is dependent on the quality and completeness of the copy made available to ProQuest.



Distributed by  
ProQuest LLC a part of Clarivate ( 2024).  
Copyright of the Dissertation is held by the Author unless otherwise noted.

This work is protected against unauthorized copying under Title 17,  
United States Code and other applicable copyright laws.

This work may be used in accordance with the terms of the Creative Commons license  
or other rights statement, as indicated in the copyright statement or in the metadata  
associated with this work. Unless otherwise specified in the copyright statement  
or the metadata, all rights are reserved by the copyright holder.

ProQuest LLC  
789 East Eisenhower Parkway  
Ann Arbor, MI 48108 USA