



ULAB

UNIVERSITY OF LIBERAL ARTS
BANGLADESH

MEDIA STUDIES AND JOURNALISM DEPARTMENT

INTERNSHIP REPORT

Mashrib Mazumder

172012022

SUMMER 2025

MSJ4298: Internship

HOST ORGANIZATION



SUBMISSION DATE: 26.08.25

Table of content

Student Details.....	2
Internship Details.....	3
Acknowledgement.....	4
Plagiarism Declaration.....	5
Internship Report Approval.....	6
Self-Assessment Form.....	7-8
Introduction.....	9
Expectation/Objectives.....	10
Host Organization.....	11
Company Organogram.....	12
Work Place Environment.....	13-14
Key Task During Internship.....	15
Project 1.....	16-20
Project 2.....	21-53
Project 3.....	54-61
Project 4.....	62-65
Project 5.....	66-67
Project 6.....	68-71
Project 7.....	72-75
Project 8.....	76-78

Student Details

Name	Mashrib Mazumder
ID	172012022
Department	Media Studies and Journalism
Major Concentration	Digital Journalism
Minor Concentration	English
Career Objectives	I am Mashrib Mazumder, major in digital journalism with a strong focus on research, storytelling, and media innovation. My work explores themes of education, justice, misinformation, culture, and digital literacy, combining traditional reporting skills with modern tools like data visualization, mobile journalism, and investigative methods. As an aspiring researcher and fact-checking specialist, I aim to use journalism not only as a way to inform but also as a way to create awareness, preserve stories, and inspire positive change. Through my projects, I have developed skills in media analysis, visual storytelling, and audience engagement, while staying committed to truth, ethics, and social responsibility.

Internship Details

Host Organization	688 Beribadh Road Mohammadpur Dhaka - 1207, Bangladesh
My Working Session	Research and Fact-Checking
Supervisor	Shuvashis Dip, Assistant Editor
Designation	Intern
Start Date	May 15th, 2025
End Date	August 15th, 2025
Hours Per Week	Around 35 hrs
Total Working Period	16 Weeks

Acknowledgement

I would like to express my sincere gratitude to Mr. Shuvashish Dip for first connecting me with CQS during my student years and for encouraging me to pursue my research idea on Media Literacy in the Age of Artificial Intelligence. His guidance in discussing theoretical frameworks and relevant literature has been instrumental in shaping my academic interests and laying the foundation for this work.

Following the completion of my studies, I was given the opportunity to join CQS as an intern. From the very beginning, I was engaged in a research project under the supervision of Mr. Naeemul Hassan, Ph.D., Associate Professor, University of Maryland, College Park. His supervision and feedback significantly enhanced my understanding of research design, methodology, and data analysis.

I am also grateful to Dr. Sumon Rahman for considering me for involvement in this important research initiative. In addition, I would like to acknowledge my colleague Mr. Ishmam Rahim, whose consistent advice and practical insights were of great value whenever I encountered challenges during my work. His collegial support made my internship experience more productive and enriching.

This internship at CQS has provided me with meaningful academic exposure and professional development, and I remain thankful to all who supported and guided me throughout this journey



PLAGIARISM DECLARATION

1. I know that plagiarism means taking and using the ideas, writings, works or inventions of another as if they were one's own. I know that plagiarism not only includes verbatim copying, but also the extensive use of another person's ideas without proper acknowledgement (which includes the proper use of quotation marks). I know that plagiarism covers this sort of use of material found in textual sources and from the Internet.
2. I acknowledge and understand that plagiarism is wrong.
3. I understand that my research must be accurately referenced.
4. This assignment is my own work, or my group's own unique group assignment. I acknowledge that copying someone else's assignment, or part of it, is wrong, and that submitting identical work to others constitutes a form of plagiarism.
5. I have not allowed, nor will I in the future allow, anyone to copy my work with the intention of passing it off as their own work.

Name: Mashrib Mazumder

ULAB ID: 172012022

Signature Mashrib



INTERNSHIP REPORT APPROVAL FORM

Student Name Mashrib Mazumder ID No. 172012022 Major: Digital Journalism Year: 2025

Local Phone: 01521309996 Email: mashrib.mazumder.msgj@ulab.edu.bd

INTERNSHIP/PROJECT INFORMATION

Internship Organization/Project: Center for Critical and Qualitative Studies (CQS)

Internship /Project Report Topic: _____ Internship Report _____
Semester (Tick one): Spring 2025 ☐ **Summer 2025** ☐ Fall 2025 _____
----- Advisor APPROVAL

(Please place a Tick)

Approved ☐ Not Approved ☐ Partial Re-write ☐ Full Re-write Remarks (If Not Approved/Pretrial/Full Rewrite):

Signature: _____ Date: 26/08/25

Name: Syeda Sadia Mehjabin Title: Senior Lecturer

SELF-ASSESSMENT FORM

The purpose of this survey is to obtain students input on how they think they have performed.

Name of Intern: __Mashrib Mazumder_____

University ID: __172012022_____

Name of the Institute/Organization:	Center for Critical Qualitative Studies (CQS)
Corporate Office Address:	House No 84, Nanda Kanan, Road 7/A, Dhanmondi - 1209
Email: cqs@ulab.edu.bd	Web: www.cqs.ulab.edu.bd/

1. During a typical week, approximately how many hours did you dedicate to your internship?

35 Hours/ Week

2. Was your internship paid? If yes, how much?

5,000/=

3. Was your site supervisor available to meet when needed? Tick any one of the following.

☐ Not at all
 ☐ Slightly
 ☐ Moderately
 ☐ Quite a bit
 ☐ **A great deal**

4. Next, we would like to know how you were affected by your internship. Using the scale below, please indicate the degree to which participation in your internship experience increased or strengthened your abilities, attitudes, awareness or understanding in each of the areas identified:

1=Not at all 2=Slightly 3=Moderately 4=Quite a bit 5=A great deal

NO	CRITERIA (Personal)	RATING
1	I have a sense of satisfaction in doing something worthwhile.	5
2	I believe in my ability to make a difference.	4
3	I am open to new experiences.	5
4	I have gained the capacity to be more productive.	4
5	I can recognize my personal strengths.	5
6	I can recognize my personal weaknesses.	4
7	I have a sense of personal achievement.	4
8	I have the ability to persevere in difficult tasks	5
NO	CRITERIA (Social/Interpersonal Impact)	
9	I demonstrate concern for the welfare of others	5
10	I can be understanding and appreciative of people with diverse backgrounds	5
11	I have the ability to work cooperatively with others	5
12	I have the ability to communicate effectively	5

13	I have increased my ability to be a leader	4
14	I feel more connected to my community	5
NO	CRITERIA (Academic learning impact)	
15	I have acquired knowledge from the job duties I performed	5
16	I have broadened my critical thinking skills (reasoning, problem solving)	5
17	I have learnt the ability to work independently	5
18	I feel my classroom learning was enriched	5
19	I have the ability to connect academic subject matter to the real world.	5
NO	CRITERIA (EMPLOYMENT/JOB SPECIFIC IMPACT)	
20	I developed specialised technical skills for a specific job function	4
21	I broadened my future employment possibilities	5
22	I received an opportunity to explore a specific career	5
23	I developed realistic ideas about the work world	5
24	I narrowed my future possible career choices.	5
25	What is your average score from this test?	110

5. General Comments

I used to think research was something only elite students pursued, but that changed when I began my internship at CQS. Thanks to my prior research experience and genuine interest, I got involved in a project from day one. This research is supervised by an assistant professor from the University of Maryland. Along the way, I faced a few dead ends, but with guidance from both my organizational supervisor and research supervisor, I was able to navigate through them. Now, I'm planning to build a career in research, as I believe I can contribute to new knowledge and explore fresh ideas.



Signature of Intern

Introduction

During my internship at CQS, I had the opportunity to engage in research projects that explored critical issues at the intersection of media, technology, and society. My work primarily involved three major data collection initiatives: fact-checking articles on misinformation using AI, analyzing AI-generated deepfakes circulating on social media, and examining gender-based misinformation. In each of these projects, I was extensively involved in collecting and categorizing data, often working independently to ensure a comprehensive and reliable dataset. Alongside data collection, I conducted literature reviews to identify relevant academic works that aligned with our research focus and contributed to designing research frameworks informed by prior studies. I also carried out pilot testing to assess the validity and effectiveness of these frameworks before full-scale implementation.

To ensure accuracy and consistency, our team adopted an odd-number cross-reviewer method, where categorizations were cleaned, verified, and reviewed collaboratively. This process provided me with valuable insight into how large-scale datasets are curated and validated for research purposes.

Beyond research, I actively participated in workshops both within CQS and at partner institutions such as BUP and ULAB. I was often responsible for preparing presentation slides and assisting my supervisor during these sessions. I also contributed to a TGI-led workshop, where I supported students in using digital tools and provided logistical assistance to the invited speakers.

This internship offered me hands-on experience in the practical dimensions of research, data management, and knowledge-sharing. It not only strengthened my technical and analytical skills but also deepened my understanding of how collaborative academic research can address pressing challenges such as misinformation, AI ethics, and digital literacy in today's media landscape.

Expectations/Objectives

The primary goal of my internship at CQS was to gain practical research experience and apply my academic training in a professional setting focused on misinformation, AI, and digital literacy. I expected to deepen my understanding of how data-driven research is conducted, strengthen my ability to design and test research frameworks, and contribute meaningfully to projects addressing critical issues such as fact-checking, deepfakes, and gender-based misinformation. I also sought to enhance my skills in literature review, data categorization, and cross-verification methods, while developing greater confidence in presenting and assisting during academic workshops.

Through this internship, I aimed not only to improve my technical and analytical abilities but also to experience the collaborative nature of professional research, where teamwork, critical thinking, and communication are essential. I was eager to learn how academic research can be translated into practical solutions for combating misinformation and to gain insight into the organizational structure of a research institute like CQS.

- Understand the research process at CQS, including data collection, categorization, and validation.
- Conduct literature reviews and contribute to designing research frameworks based on previous studies.
- Gain hands-on experience in pilot testing and applying verification methods such as cross-reviewer categorization.
- Assist in workshops by preparing materials, supporting supervisors, and engaging with participants.
- Strengthen research, analytical, teamwork, and communication skills through practical involvement in ongoing projects.
- Explore the applications of AI in fact-checking, misinformation detection, and digital literacy research

Host Organization

The Center for Critical and Qualitative Studies (CQS) is a leading research center under the University of Liberal Arts Bangladesh (ULAB), dedicated to advancing interdisciplinary research and fostering a culture of critical inquiry. Since its inception, CQS has positioned itself as a hub for innovative scholarship, with a particular emphasis on media studies, communication, cultural analysis, and the social sciences. The center promotes rigorous academic research while bridging the gap between theoretical knowledge and real-world application, especially in areas that shape contemporary society.

CQS is well-recognized for its focus on qualitative methodologies, critical theory, and interpretive approaches, which allow for nuanced understanding of complex social phenomena. Its projects often explore issues such as misinformation, media literacy, artificial intelligence in communication, gender representation, migration, and cultural transformations in South Asia and beyond. By creating collaborative platforms that connect faculty members, students, and external researchers, CQS actively encourages knowledge-sharing, critical debate, and interdisciplinary exploration.

In addition to research, CQS is committed to building academic capacity through workshops, seminars, conferences, and publications. These initiatives are designed to train the next generation of scholars in both theoretical and applied aspects of research. The center regularly organizes collaborative events with local and international institutions, bringing together academics, practitioners, and students to discuss pressing global and regional challenges. Its work has contributed to building stronger linkages between academia and society, ensuring that research outcomes are not only academically robust but also socially impactful.

CQS also plays a key role in mentorship and training, offering opportunities for students to participate in ongoing projects, engage in data collection and analysis, and gain exposure to professional research practices. By involving young scholars in areas such as fact-checking, digital literacy, and the study of AI-generated content, the center ensures that research remains contemporary, relevant, and responsive to emerging global trends.

As part of ULAB's broader vision of providing world-class liberal arts education, CQS contributes significantly to the university's reputation as a space for critical thinking, creativity, and social responsibility. Its dedication to qualitative and critical research reflects ULAB's commitment to nurturing informed citizens and researchers who can engage meaningfully with the challenges of the 21st century.

Company Organogram



Workplace Environment





Key Tasks During the Internship

Task 1: Data Collection and Categorization

- Conduct extensive data collection on three major areas: fact-checking articles on misinformation using AI, AI-generated deepfakes on social media, and gender-based misinformation.
- Single-handedly categorize and organize datasets to ensure structure and usability.
- Perform pilot testing of research frameworks to evaluate effectiveness and refine approaches.

Task 2: Literature Review and Framework Design

- Conduct literature reviews to identify relevant theories and prior studies that align with ongoing research.
- Assist in designing research frameworks grounded in previous scholarship.
- Support the validation of datasets through an **odd-number cross-reviewer method** to ensure accuracy, consistency, and reliability.

Task 3: Workshop Participation and Support

- Participate in internal workshops at CQS and external sessions at BUP and ULAB.
- Prepare presentation slides and provide assistance to supervisors during academic workshops.
- Contribute to a **TGI-led workshop** by helping students navigate digital tools and supporting guest speakers in delivering their sessions.

Project 1

Project Name: Hyper-realistic Deepfake Contents in Social Media

Start Date: 15th May 2025

End Date: To be Continued

Project Overview:

This project investigates the rise of AI-generated misinformation and deepfake content, with a special focus on Facebook as a platform where manipulated media circulates widely. The research aims to understand how deepfakes are created, shared, fact-checked, and received by audiences. By building a dataset of over 1,000 deepfake items (600+ videos with ~340 million views) and cross-referencing these with fact-checking reports, the study identifies gaps in detection, debunking, and narrative framing of misinformation.

My Tasks and Responsibilities:

In this project, my main responsibilities included compiling and analyzing a dataset of over 1,000 deepfake items, categorizing them by technique and narrative, and cross-checking them with fact-checking reports to identify coverage gaps. I examined selected cases to study audience responses and engagement patterns. Additionally, I contributed to designing survey and interview tools and conducted a literature review to frame the study within global research on misinformation and fact-checking practices.

My Data Collection of 1000 Deepfake Items from Social Media :

Dataset Development & Analysis

- Compiled a dataset of **1,028 deepfake items**, categorized by technique (face-swapping, lip-syncing, puppet motion, edited renderings, manipulated reality) and narrative (satirical parody, fake political statements, propaganda, misleading content).
- Cross-checked against **109 fact-checking reports** to identify coverage gaps and reasons why viral posts often remain unverified.
- Conducted in-depth analysis of **30 selected cases**, focusing on audience engagement and discourse in comment sections.

Literature Review & Theoretical Framing

- Conducted a structured review of **scholarly and industry literature** on deepfakes, misinformation, and fact-checking practices across multiple countries.

Documentation & Knowledge Contribution

- Synthesized findings into project documentation highlighting observed patterns, research gaps, and theoretical contributions.
- Identified **emerging global patterns** of AI misinformation and fact-checking, connecting Bangladesh's context with international debates on deepfakes and media literacy.

Dataset : [Hyper-realistic Deepfake Contents in Social Media](#)

A	B	C	D	E	F	G	H
Post Link	Post Text	Reaction	Comment	Share	Views	Page/Group/Profile Name	Media Type
https://www.facebook.com/watch/?v=69d-69c25B&as=12862015472&fbid=9299dcf8-f3d0-gy3DEVBJ	সবার থেকে সন্তুষ্ট সত্যকে: শেষ হারিয়ে শাখির বিন ফুরান, এবার বিবাহের খবর?	703	29	160	19k	Barnister Zaima Rahman	AI Generated Video
https://www.facebook.com/watch/?v=6e0d-wed08xsc=190724357658107&fbid=f3-d0guyPqubJal	ছাত্র অনবার টিউশন শেষে হাসিনিয়ার বাসা থেকে কতকাল সে সভা থকা বলে দেখাবে		1100	90070	1.5m	দাশপাক আল আমিন	Lip Sync/Manipulated Audio
https://www.facebook.com/watch/?v=okwzweo103t113080960	শেষ হাসিনি নই, আসল হাসার অব্যক্তিক হিসাব আমি কেন জানেন? আমি কোন্ট্রোলিং, প্রকাশনাওঁ পাইনি, নিজে নিজেকে মিটার আমার একটা-একটি দিলে... কখনো কোনোদিন না, যোগ করে তিনি শেষ হাসিনি আরও কিছু, আমি যদি শুধি নিয়ে মারা মুখে, কেউ আমাকে কাছে ছুঁতামি তবে বলি তোলা। শেষ কোনও না থাকবে।	58300	3300	58900	6.2m	RTNews24	Lip Sync/Manipulated Audio
https://www.facebook.com/watch/?v=3170247932040412	ডঃ সুবির জগৎ বিশ্বেষ কেনে বেশ ভালোভাবে, মেলে জায়, ভারে টিউনিং ও অফারহীলিশের দুখীনকার কোচেরে লারী করেন শেষ হাসিনি মেথৌকে ধাক দিয়ে পাঠি ছাড়তে কলেক্স ট্রাম্প[Donald Trump] AI Creation	13000	2000	4500	415.5k	RTNews24	Lip Sync/Manipulated Audio
https://www.facebook.com/news24.bd/videos/598929219493328	যুক্তিবিরতি দেখাতে যে পার বিরোধের রোনাস্ক ট্রাম্প বালাদেশের গেসিটেট হতে সন রোনাস্ক ট্রাম্প Donald trump] AI Creation	31000	426	18000	1.6m	RTNews24	Lip Sync/Manipulated Audio
https://www.facebook.com/news24.bd/videos/525691736491590	শেষ হাসিনি শিয়র ৪০০ কোটির ছড়িক, আর আমি গেসিটেট হয়েছেন তৈরি ভিকা করি, বাংলাদেশের গেসিটেট না হয়ে শেষ হাসিনির অনেক	3600	122	2700	163k	RTNews24	Lip Sync/Manipulated Audio
https://www.facebook.com/news24.bd/videos/723695190119563	নেত্রুর প্রয়োজন মানুষের চেয়েও কম সুখে নেওয়ার জন্য, ডয়ের পৃথক করার জন্য বোভল চুড়ে মারার ঘটনার মধ্যস্থতকে কি বুঝি দিনের মধ্যে মার্চে?	265	4	143	15k	RTNews24	Lip Sync/Manipulated Audio
https://www.facebook.com/news24.bd/videos/483436594791964	"মীর কাশেম খাদী ঘনি আমেরিকা জায়েদের, জর্জ রাশিসনের পক্ষে... See more	747	23	618	34k	RTNews24	Lip Sync/Manipulated Audio

Findings:

This project investigates the role of deepfakes as a dominant form of political commentary in Bangladesh's digital ecosystem, with a particular focus on their circulation, emotional impact, and narrative framing. The analysis of 1,028 AI-generated contents, including 600 videos with nearly 340 million collective views, demonstrates that deepfakes have become central to shaping political discourse online. These videos are not merely entertainment; rather, they function as tools of persuasion, satire, and disinformation, targeting both leaders and activists in ways that amplify existing political divides.

A key finding is that the majority of analyzed deepfakes fall under the category of political satire, designed to provoke strong emotional reactions such as humor, mockery, fear, and nationalism. This pattern intensified after August 5th, when Sheikh Hasina was first depicted as a deepfake character. The introduction of her likeness marked a turning point, triggering unprecedented public curiosity, virality, and demand for such content. Sheikh Hasina remains the most frequently featured figure, often shown through highly realistic AI-generated visuals. These portrayals revolve around post-resignation narratives, loss of political control, and derisive mockery, resulting in massive audience engagement through views, comments, and shares.

Beyond Hasina, Narendra Modi frequently appears in deepfakes with an anti-Indian stance, employing provocative tones that stir nationalistic sentiment within Bangladeshi audiences. Meanwhile, Dr. Muhammad Yunus is targeted in content circulated by pro-Hasina networks, where he is framed as incompetent in relation to the interim government—acting as a counter-narrative to pro-opposition discourse. Furthermore, female activists and celebrities, such as Farzana Sithi and Umama from the July protests, are recurrent subjects of gender-targeted deepfakes. These often use face- or body-swap technologies to ridicule or discredit their activism, underscoring a disturbing trend of digitally gendered political violence.

Despite the scale of this phenomenon, only 120 items out of 1,028 were fact-checked by third-party organizations such as Rumor Scanner, Boom Bangladesh, and Fact-Watch. This indicates a major gap in verification, as most viral or emotionally charged posts remain unchallenged. The majority of deepfakes originate from Facebook pages, while individual profiles contribute relatively little, suggesting that organized networks, rather than lone actors, play a larger role in their distribution.

The findings highlight both the manipulative intent and psycho-social consequences of these contents. Political satire, when delivered through synthetic media, can desensitize audiences to the seriousness of political leadership, normalize misinformation, and pave the way for more harmful uses such as fake speeches, altered evidence, and character assassinations disguised as humor. The study also identifies a troubling feedback loop between AI-generated content and platform algorithms: during data collection,

repeated interactions caused Facebook's algorithm to amplify and recommend more political satire deepfakes, reinforcing exposure and creating a cycle that sustains their visibility and influence.

Finally, while current deepfake tools still display imperfections such as audio mismatches and visual anomalies, the technology is rapidly advancing. Newer models (2025+) show greater accuracy in lip-syncing, emotional tone, and high-resolution rendering. This increasing realism enhances the persuasive power of deepfakes and strengthens their role as instruments of political propaganda

Limitations and How I Overcame Them

- **Uncertainty in Categorization Framework**

Limitation: At the start of data collection, the category framework for classifying deepfakes was still undecided. This created ambiguity in how to systematically organize and analyze the dataset.

- ***Overcome:*** We began with an initial framework based on our team's understanding and ongoing observations. As the literature review progressed in parallel, we refined and finalized a robust categorization framework that was grounded in published academic research.

- **Simultaneous Data Collection and Literature Review**

- ***Limitation:*** Data collection had already started before a fully validated framework was in place, which risked inconsistency in classification and analysis.
- ***Overcome:*** By conducting the literature review alongside data collection, we continuously updated and adjusted the framework. This ensured that the dataset was eventually aligned with theoretically sound categories supported by established studies.

- **Need for Validation of Framework**

- ***Limitation:*** The initial framework was not formally tested, raising concerns about reliability and applicability across different types of content.
- ***Overcome:*** A pilot test of the framework was carried out under the supervision of my project supervisor, which confirmed its effectiveness. The dataset was then refined accordingly, making it functional and consistent with frameworks published in peer-reviewed literature.

Project 2

Project Name: Understanding Deepfake Misinformation and Fact-Checking in Social Media

Start Date: May 20, 2025

End Date: Jul 31, 2025

Project Overview:

This project focused on a dataset to analyze misinformation and synthetic media generated through artificial intelligence. The primary aim was to identify, categorize, and assess deepfakes and manipulated content circulating on Bangladeshi social media platforms. Using fact-checking sources such as Fact-Watch, Rumor-scanner, Boom-Bangladesh, Dissmilab, I collected and organized articles, posts, and multimedia content that exhibited signs of AI-driven manipulation. The dataset captured details including post text, engagement metrics (reactions, comments, shares, and views), and page/group origins, along with classification of manipulation techniques, influence types, and debunking methods.

My Tasks and Responsibilities:

In this project, I was primarily responsible for collecting and categorizing posts related to AI-generated misinformation, including deepfakes and manipulated content. I conducted literature reviews to identify relevant theories and frameworks, and helped in designing survey questions for fact-checkers and content sharers. I also carried out two pilot tests of the framework, analyzed challenges in classification, and contributed to refining the model by integrating additional variables from global studies. To ensure dataset accuracy, I participated in the three cross-reviewer validation process, which improved the reliability and consistency of our categorizations.

Data Collection: Posts were collected between August 2024 and 31 July 2025, focusing on AI-generated misinformation and manipulated content. These posts were systematically catalogued to trace patterns of dissemination and audience engagement.

Fact-Checking Cross-Reference: Posts containing fact-checking articles were identified and marked, enabling a comparison between misinformation flows and the corrective efforts made by fact-checking organizations.

Literature Review: Relevant scholarly and industry literature with similar categorization methods was reviewed, ensuring that the project's framework is aligned with international standards and comparable studies.

Framework Development: Based on the literature, survey questions were developed for fact-checkers as well as individuals who engage in sharing misinformation online.

Pilot Testing: Two pilot tests of the framework were conducted.

The first pilot revealed that some posts required modifications within the framework

due to the dual nature of AI-generated misinformation, which is both manipulative and misleading.

To address this, we incorporated insights from other studies that directly examine this complexity.

Key Concepts Identified:

AI Deception – creates false beliefs.

AI Manipulation – aims to influence attitudes, emotions, or decisions.

Framework Integration: Another misinformation framework outlining eight distinct variables was reviewed. We merged its structure with our own, adding extra elements such as engagement metrics and narrative framing to strengthen analytical depth.

Cross-Reviewer Validation: To improve reliability and consistency, a three cross-reviewer method was implemented. This ensures cleaner categorization, minimizes individual bias, and provides a more robust dataset.

Dataset : [Understanding Deepfake Misinformation and Fact-Checking in Social Media](#)

AI and Misinformation and Fact-Checking						
File Edit View Insert Format Data Tools Extensions Help						
100% 123 Arial 11 B I A						
t1000						
	A	B	C	D	E	F
		Article Link	Post Link	Post Text	Reaction	Comment
1						
2	Fact-Watch	https://www.fact-watch.org/political-deepfakes-election-campaigns/	https://www.facebook.com/watch/?mibextid=wwXfr&y=1607251676681074&rcid=FcD0u8gPdq3uFxd	ছাত্র জনতার উদ্দেশ্যে শেখ হাসিনার ভাষণ মোক কসকে সব সত্য কথা বলে দেওয়া		1100
3	Fact-Watch	https://www.fact-watch.org/political-deepfakes-election-campaigns/	https://www.facebook.com/watch/?v=1787257932040412	ভরা ভূমির জন্য নিম্নের বোন শেখ বেহানা, হেলে জয়, ভায়ে টিউলিপ ও আওয়ামীলীগের মুশীতিবাজ নেতাদের দাবী করেন শেখ হাসিনা বাংলাদেশের জেনিটেক হতে চান ডোনাল্ড ট্রাম্প। Donald trump। AI Creation	13000	2000
4	Fact-Watch	https://www.fact-watch.org/political-deepfakes-election-campaigns/	https://www.facebook.com/news24.bd/videos/525591280491080	শেখ হাসিনার শিরন ৪০০ শ্রেণীর মালিক, আর আমি জেনিটেক হতেও স্নেন ডিকা করি, বাংলাদেশের জেনিটেক না হতে গেলে ট্রাম্পের আবেগ	3600	122
5	Fact-Watch	https://www.fact-watch.org/political-deepfakes-election-campaigns/	https://www.facebook.com/news24.bd/videos/525591280491080	আমাদের মনে মনে মূল্যবান শেখ হাসিনা হাসিনার কাছাকাছি থাকতে—কিভাবে আমাদের খেঁচি হতে গিটীংটী সম্পূর্ণ কাল্পনিক, শুধু মাত্র আমাদের ভাবনাতে তুলে ধরতে আটাইনশিয়াল প্রকৃতির মাধ্যমে তৈরী করা হয়েছে।	61100	4600
6	Fact-Watch	https://www.fact-watch.org/political-deepfakes-election-campaigns/	https://www.facebook.com/news24.bd/videos/525591280491080	এই আদর্শ পালিয়ে নিয়েছিলাম নাকি তাকিয়ে নিয়েছিলাম? মূল্যবান শেখ হাসিনা ছাত্র জনতার উদ্দেশ্যে শেখ হাসিনার ভাষণ	119000	15500
7	Fact-Watch	https://www.fact-watch.org/two-ai-photos/	https://www.facebook.com/photo/?id=558856140204825&set=a.157499317007379	অসম্মিত ছবি তুলির নিম্নে মার্কট্টে চুক ছবি মুঠি বাওবেশ	948	37
8	Rumor Scanner	https://rumor-scanner.com/fact-check/indian-vlogger-photo-claiming-ai-tress-sadia-ayman-have-been-ai-created/149796	https://www.facebook.com/story.php?story_fbid=122172175759632958&id=81559887586748&rcid=jq1BK1pyu63ao#	ছাত্র কপ সাদিয়া আহমান	5100	1500
9	Rumor Scanner	https://rumor-scanner.com/fact-check/ai-generated-image-on-fake-claim-pregnant-police-died-in-quota-protest/147279	https://www.facebook.com/100003035924036/posts/9401549619956242/?id=GEKRISQmutk1v	অসম্মিত ছবি তুলির নিম্নে মার্কট্টে চুক ছবি মুঠি বাওবেশ এই নারকীয় হত্যার বিষয়ে এই মাস্টে কি হবে? এই পেন্সী কি মানুষের কনবাস যোগ আছে? অন্যের শিশুটি মারের শুলি বরং খাম নিয়ে পুঁথির আগে পৌঁছায় আগেই তাকে করে বেত হলেই। হায় মানুষ! মানুষ মারের শাসনকে কনের শৈশবিকতার পেটা পেলে আর কনবাসের অসম্মিত ছবি তুলির নিম্নে মার্কট্টে চুক ছবি মুঠি বাওবেশ। আগেই পুঁথির উপরে খট খটয়া খটন সত্তা! তুমি ছাড়া কেবল নিজে খট খটয়া সত্তাকে ধাপ সোয়া ককী যৌতিক? যদি বলি কটীয়াত তুমি; তবে কি সিরাজগঞ্জের	307	31
Sum: 1,799,933.00						

Version 1:

[illegible]

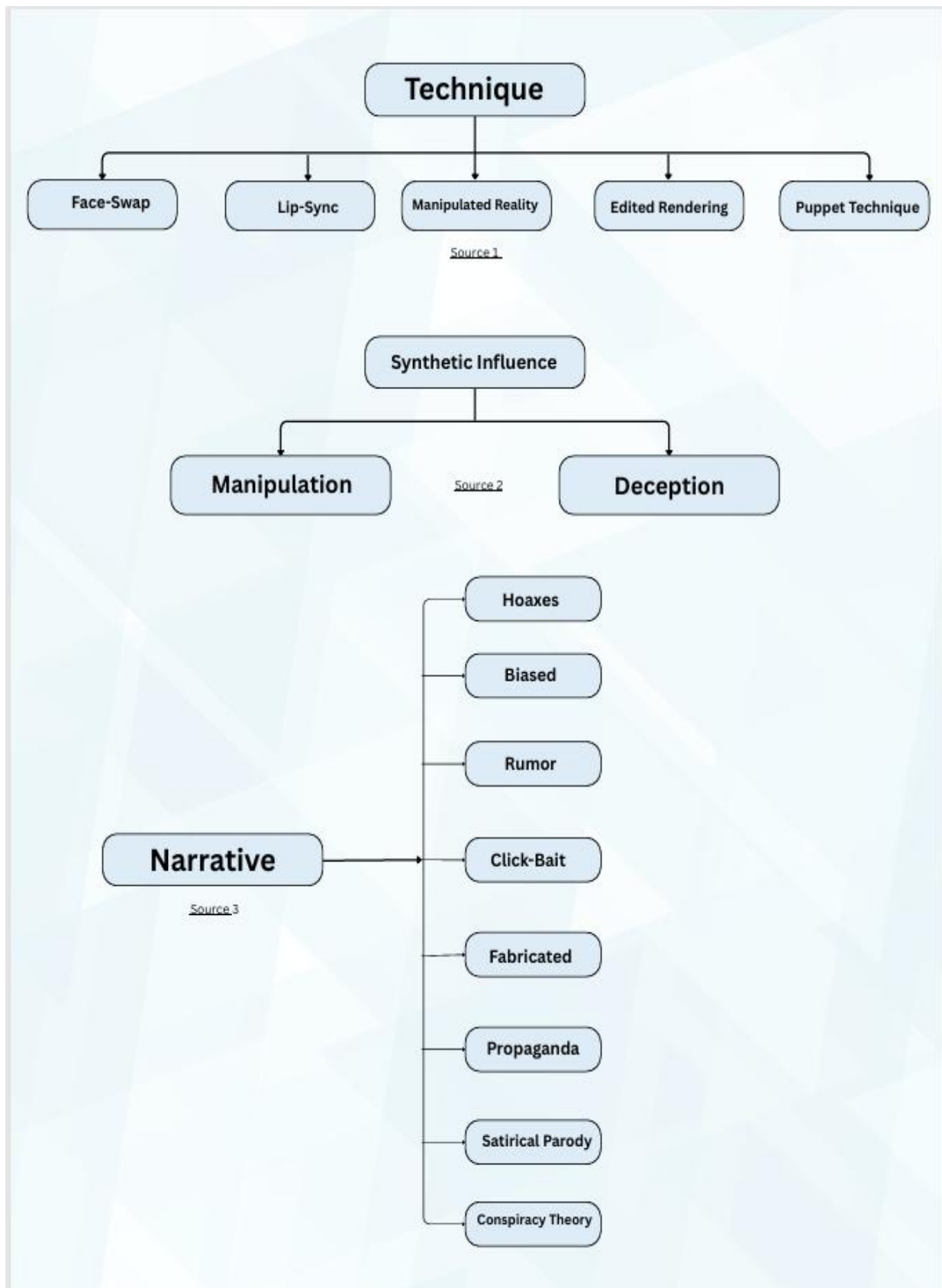
Version 2 :

[illegible]

Narrative Description:

A	B	C	D	E
Fabricated (F)				
Completely made-up stories with no truth.				
Example: A story that Hillary Clinton adopted an alien baby.				
Propaganda (P)				
Made-up stories with political motives to harm a person or group.				
Example: Fake stories during elections or wars to sway opinions.				
Conspiracy Theories (CT)				
Claims of secret plots without evidence.				
Example: Pizzagate or false claims about Seth Rich's murder.				
Hoaxes (H)				
False stories presented as truth, often with some real details.				
Example: Fake celebrity death news.				
Biased or One-Sided (B)				
Extremely one-sided stories, often political.				
Example: Hyperpartisan news from alt-right communities.				
Rumors (R)				
Unverified or unclear stories spread widely.				
Example: False claims after the Boston Marathon bombing.				
Clickbait (CL)				
Misleading headlines used to attract clicks.				
Example: "You won't believe what happened next..."				
Satire (S)				
Funny or ironic fake news, not meant to mislead—but often misunderstood.				
Example: Articles from The Onion mistaken as real.				

Proposed Frame Work:



Research Questions from previous research

RQ1. What dangers do fact-checkers identify in connection with deepfakes for (1) journalism and (2) news audiences?

RQ2. To what extent do fact-checkers adapt their work routines to deepfakes in terms of (1) detection techniques, and (2) countermeasures?

RQ1. What are the key trends in the creation and dissemination of AI-generated misinformation, including topics, origins, targets, intentions, elements, and generative models in Brazil, Germany, and the UK?

RQ2. What detection methods and debunking strategies are employed by fact-checkers, and what additional insights do they provide for countering AI-generated misinformation in Brazil, Germany, and the UK?

RQ3. How do fact-checkers assess the ongoing spread of AI-generated misinformation, and how do they address the challenges it poses?

RQ1

What is the frequency of debunking compared to fact-checking articles during the Brazilian election run-off period?

Due to the influence of the 3PFC program, Brazil's low resilience to online disinformation, and its high social media use, the quantity of debunking articles may surpass the number of fact-checking articles.

RQ2

Which actors did Brazilian fact-checkers monitor during the electoral run-off period?

Despite their commitment to the norm of balance, we assume that fact-checkers will focus more on far-right political agents since they are more likely to use disinformation as a political communication strategy.

Bearing in mind that fact-checkers' performance is built upon traditional journalistic values, such as prominence (of disinformation agents and targets), we also ask:

RQ3

Who are the victims of disinformation, that is, the target actors portrayed in the verification articles?

Considering the strategic use of disinformation by Jair Bolsonaro's 2018 campaign, we expect that fact-checkers will focus on verifying a greater number of falsehoods targeting the opposition and the media, who are the usual targets of Bolsonaro's fake news machinery.

RQ4

Which social media platforms were primarily monitored during the 2022 Brazilian presidential election?

Based on the data provided by the 2022 digital news report and Meta 3PFC, fact-checkers will likely prioritize Facebook and WhatsApp, the most used social media platforms for news in the country.

RQ5

Which topics and types of falsehoods were primarily scrutinized by Brazilian organizations?

Due to the lack of studies or different research designs regarding fact-checkers' selection of falsehood types, especially in the Global South context, we kept RQ5 open.

RQ1. Are there differences in the behavioral intentions of different groups (e.g., males and females, youth, and adults) towards deepfake videos?

Following the findings from RQ1, we also investigate how the audiences' behavioral tendencies (e.g., watching or sharing behavior) might influence the deepfake effects. It might reasonably be assumed that if a

deepfake is not spread or believed by its audience, any negative impact will be small, but this inference remains largely unstudied. Therefore, RQ2 is proposed.

RQ2. Are there any direct and/or indirect relations between participants' behavioral tendencies of watching or sharing deepfakes, and the consequences and impacts of deepfakes on the target or society?

RQ3. Do the content and the production quality of deepfakes influence the audience's behavioral intentions in terms of watching and/or sharing deepfakes?

RQ1: What is the difference in the perceived credibility and believability between factually accurate information and disinformation across modalities?

RQ2: What is the difference in the perceived credibility and believability between deepfakes and cheapfakes?

RQ1. What kinds of fake content are developed by AI around political leaders on TikTok?

RQ2. How much do deepfakes change people's beliefs about political leaders, the media, and various institutions?

RQ3. What moral questions come up when deep-fake videos are viral against popular politicians?

RQ1. How social and political conditions make Pakistani society vulnerable to deepfake influence

RQ2. How media consumption habits contribute to the acceptance and spread of manipulated digital content?

RQ3. How social media echo chambers shape public susceptibility to deepfake-driven misinformation.

RQ1: Do new media-literate and non-media-literate people differ in their capacity to identify deepfakes?

RQ2: To what extent does the witnessed exposure of deepfake videos affect the trustworthiness of media-literate and non-media-literate people in online videos?

Literature Review

1. [Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential](#)

What They Categorized	How
Participants	By role: open-source, academic, start-up, corporate
Values	Consent, accountability, harm, privacy, transparency
Governance Themes	Design-level, institutional, and policy-level interventions
Motivations	Economic incentives, research culture, moral agency

They conducted a thematic analysis using predefined ethical values (sometimes called a "codebook" approach), such as:

1. Consent
2. Accountability
3. Transparency
4. Harm Prevention
5. Privacy
6. User autonomy
7. Authenticity

2. [Everyday encounters with deepfakes: young people's media and information literacy practices with AI-generated media](#)

Central theme	Category
Encountering deepfakes	Situation
	Platform
	Content
Responding to deepfakes	Casual interaction with deepfakes
	Action on deepfakes
Understanding deepfakes	Content creation
	Impact
	AI techniques

Twenty participants ($N = 20$) consented to be interviewed and responded to questions following a pre-prepared interview outline that included five main themes:

- AI in everyday apps
- AI and personalized content
- AI in finding media content
- synthetic media/AI-generated content
- impacts of AI on young people

The questions asked

1. Do you know about media content that is generated using AI?
2. If so, can you talk about your experiences?
3. Would you like to show an example (an app or video, etc.)?

3. [Between Realities: Information Sharing Practices of Deepfake Creators](#)

4. [You Won't Believe What They Just Said! The Effects of Political Deepfakes Embedded as Vox Populi on Social Media](#)

(Dis)information Embedding	Authentic Unrelated	Malinformation	Disinformation
No embedding	(1) Unrelated authentic speech, no embedding (an authentic fragment of a political speech on the country's progress)	(2) Decontextualized authentic speech, no embedding (the message leaves out the context of right-wing populist statements on immigration that were expressed by the political actor)	(3) Disinformation, no embedding (the depicted political actor is made to voice radical rightwing hate speech targeted at immigrants)
Endorsed		(4) The authentic message is embedded in the social media feed of an ordinary citizen who approves of the right-wing populist statements of the political actor (i.e., saying it is a strong statement)	(5) The disinformation message is embedded in the social media feed of an ordinary citizen who approves of the hate speech (i.e., saying it is a strong statement)
Discredited		(6) The authentic message is embedded in the social media feed of an ordinary citizen who disapproves of the right-wing populist statements of the political actor (i.e., saying it is a weak statement)	(7) The disinformation message is embedded in the social media feed of an ordinary citizen who disapproves of the hate speech (i.e., saying it is a weak statement)

5. [An Average Joe, a Laptop, and a Dream: Assessing the Potency of Homemade Political Deepfakes](#)

6. ["It's Up to Me Whether I Do—Or Don't—Watch Deepfakes": Deepfakes and Behavioral Intention](#)

7. [The detection of political deepfakes](#)

8. [Does trust in government moderate the perception towards deepfakes? Comparative perspectives from Asia on the risks of AI and misinformation for democracy](#)

9. [An Ecosystem Approach to Ethical AI and Data Use: Experimental Reflections](#)
10. [The ethical agency of AI developers](#)
11. [Ethics in AI through the practitioner's view: a grounded theory literature review](#)

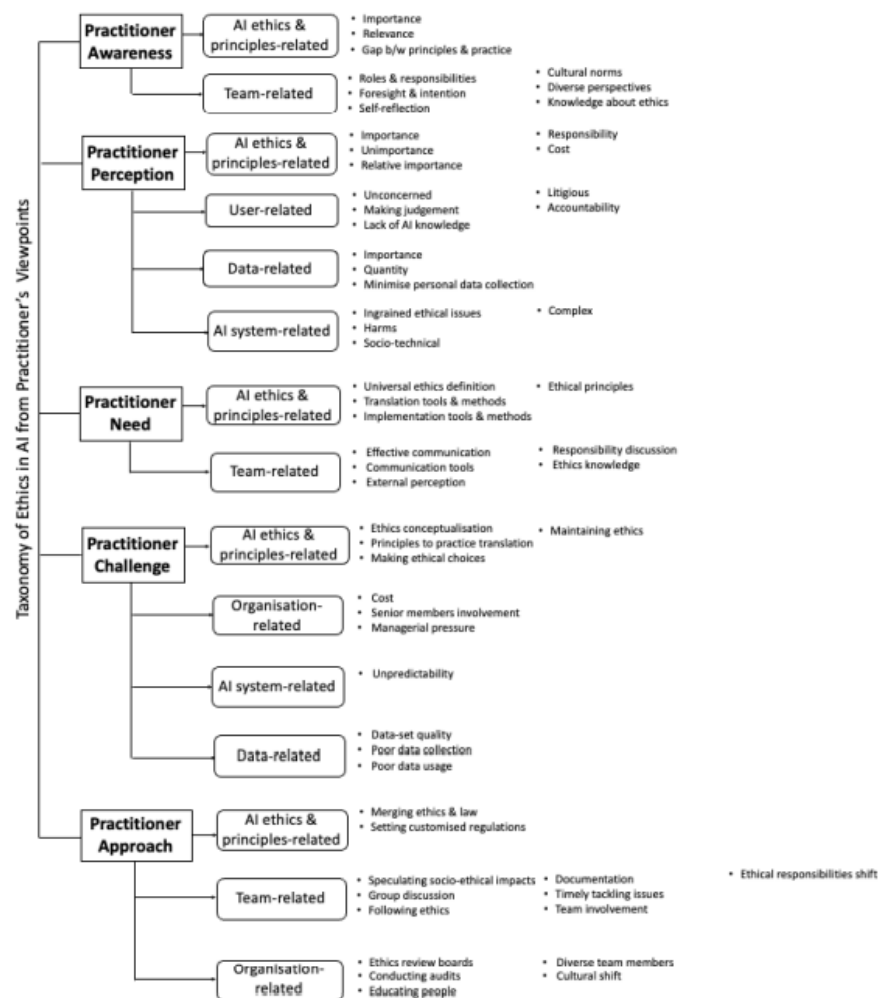


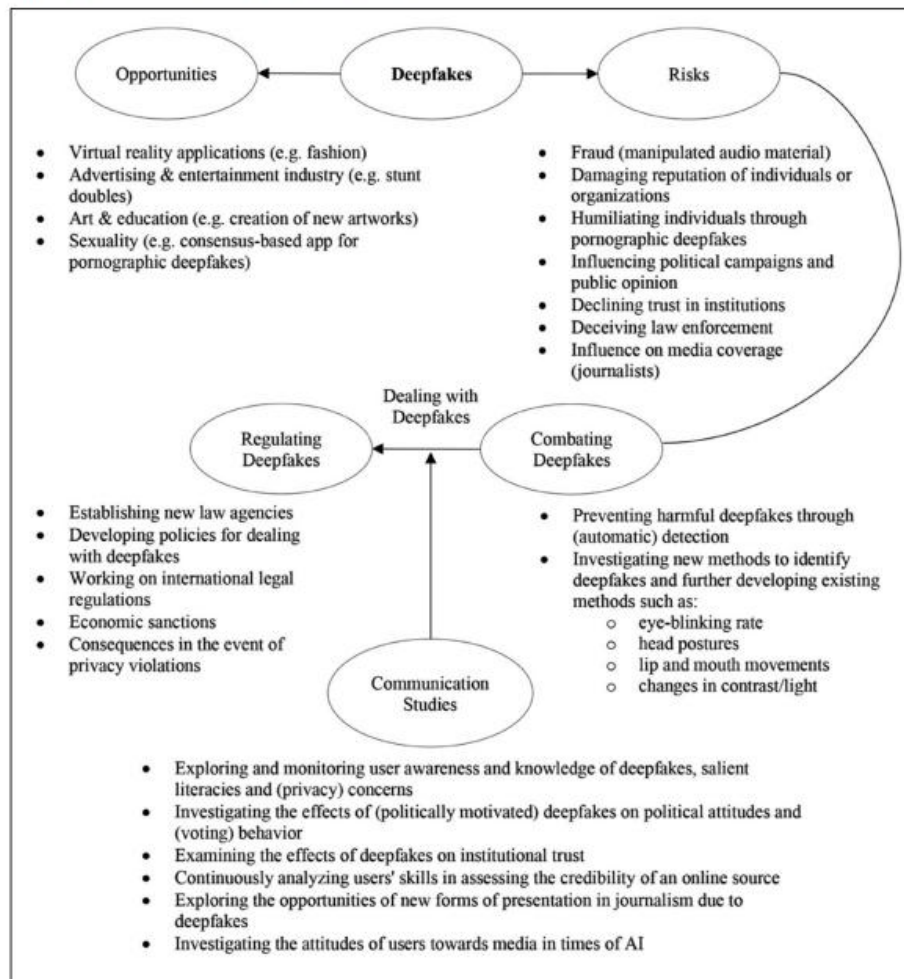
Fig. 4 Taxonomy of Ethics in AI from Practitioners' Viewpoints

12. [Ethics in AI through the Developer's Prism: A Socio-Technical Grounded Theory Literature Review and Guidelines](#)
13. [Ethics in the Age of AI: An Analysis of AI Practitioners' Awareness and Challenges](#)
14. [Cheap Versus Deep Manipulation: The Effects of Cheapfakes Versus Deepfakes in a Political Setting](#)
15. [They Would Never Say Anything Like This! Reasons To Doubt Political Deepfakes](#)
16. [Dealing with deepfakes – an interdisciplinary examination of the state of research and implications for communication studies](#)

For the search,
the following keywords were applied:

- 1) deepfake / deep fake
- 2) fake digital content
- 3) fake video
- 4) deep learning video
- 5) video manipulation.

Figure 1. Model of dealing with deepfakes



17. [Human Detection of Political Deepfakes across Transcripts, Audio, and Video](#)

Component	Description
Framework	Multimodal perception across content vs. cues; discrimination ability.
Dataset	32–60 political speeches; real vs. deepfake; TTS & voice-actor audio.
Experiments	5 pre-registered + large-scale replication; randomized modality and context.
Measured Outcomes	Detection accuracy, confidence, behavioral traces.
Analysis Tools	OLS, t-tests, multiple comparisons, response-time metrics

18. [How prevalent is AI misinformation? What our studies in India show so far](#)

19. [The origin of public concerns over AI supercharging misinformation in the 2024 U.S. presidential election](#)

1. Mixed-Methods Approach

The study uses a two-phase mixed-methods design:

Survey Experiment (Quantitative)

Conducted via the Lucid platform with a nationally representative U.S. sample (n = 1,545)

Participants were randomly assigned to one of four video conditions:

Real video (unaltered political content)

Cheapfake (manipulated timing/editing)

Deepfake (AI-altered speech/appearance)

No video (text-only control)

Open-ended Text Analysis (Qualitative)

Participants were asked to explain in their own words what concerned them most about AI and misinformation after exposure to a video or text.

These responses were then coded thematically using grounded theory-style inductive coding.

2. They build an original typology from participant responses:

Four emergent concern categories:

Authenticity Concerns – fears about fake or misleading AI content

Spread Concerns : how quickly and widely misinformation can propagate

Influence Concerns : effects on voters, especially the uninformed

Accountability Concerns : lack of clarity on who is responsible for AI content

20. [Classifying Human-Generated and AI-Generated Election Claims in Social Media](#)

21. [AI-generated misinformation in the election year 2024: measures of European Union](#)

22. [AI-Generated Misinformation: A Case Study on Emerging Trends in Fact-Checking Practices Across Brazil, Germany, and the United Kingdom](#)

Research Questions :

RQ1. What are the key trends in the creation and dissemination of AI-generated misinformation, including topics, origins, targets, intentions, elements, and generative models in Brazil, Germany, and the UK?

RQ2. What detection methods and debunking strategies are employed by fact-checkers, and what additional insights do they provide for countering AI-generated misinformation in Brazil, Germany, and the UK?

RQ3. How do fact-checkers assess the ongoing spread of AI-generated misinformation, and how do they address the challenges it poses?

Method :

Quantitative content analysis of 136 fact-check articles from 2023–2024

13 expert interviews with practitioners in AI and fact-checking

Questionnaire :

1. Experience with AI-Generated Content

How often do you encounter AI-generated misinformation?

Can you recall a recent case that stood out?

2. Detection Tools and Techniques

What tools or methods do you currently use to detect AI-generated content?

Are existing detection systems sufficient or lacking?

3. Verification & Editorial Practices

How has your editorial workflow changed due to AI-generated content?

Do you have guidelines specifically for handling deepfakes or synthetic media?

4. Cross-border Collaboration

Do you collaborate with international fact-checkers or platforms?

What are the challenges in addressing AI-generated misinformation that crosses borders?

5. Anticipation of Future Trends

What concerns you most about the future of AI and misinformation?

How do you think AI will impact public trust and fact-checking in the coming years?

23. [Mapping Global Emerging Scholarly Research and Practices of AI-supported Fact-Checking Tools in Journalism](#)

Research questions:

RQ1. How has the academic literature on algorithm-driven fact-checking tools in journalism evolved in terms of publication activity?

RQ2. Which patterns emerge in the authorship and geographical focus of research on fact-checking tools in algorithm-driven journalism?

RQ3. What are the primary AI-supported tools utilised by fact-checkers in verification procedures in the selected countries?

RQ4. Which monitoring/misinformation identification tools are currently accessible to and employed by fact-checking organisations in the selected countries?

24. ["We Follow the Disinformation": Conceptualizing and Analyzing Fact-Checking Cultures Across Countries](#)

Abstract : Democratic societies inherently depend on an informed citizenry. By shaping citizens' voting behavior, fostering political cynicism, and reducing trust in institutions, misinformation can pose significant challenges to individuals and societies. Against this backdrop, fact-checking initiatives aimed at verifying the accuracy of publicly disseminated (mis)information have flourished worldwide. However, existing research is disproportionately oriented toward the Global North, with a focus on the United States and the most influential organizations. Equally scarce are comparative studies. To address these shortcomings, this study introduces a context-sensitive framework for analyzing fact-checking cultures and illustrates its application in a cross-national comparative design by contrasting two countries from the Global South and North: Brazil and Germany. Using a mixed-methods design, we integrate computational, qualitative, and quantitative content analysis of 11 fact-checking organizations and 13,498 fact-checking articles over 11 years (2013–2023), alongside qualitative semistructured interviews with fact-checkers (N=10). Our findings reveal several areas of divergence and convergence, suggesting that fact-checking cultures transcend organizational and national boundaries.

25. [Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network](#)

26. [Adding human values on the deepfake: co-designing fact-checking solutions to combat misinformation](#)

27. [Cutting through the Hype: Understanding the Implications of Deepfakes for the Fact-Checking Actor-Network](#)

RQ1. What dangers do fact-checkers identify in connection with deepfakes for (1) journalism and (2) news audiences?

In this study, deepfakes are considered a non-human actant in the fact-checking network, where new actors are developing technology-based and manual detection techniques to uncover false and misleading visuals. In addition, fact-checkers are called to counteract digital mis- and disinformation campaigns through education and training. To understand how deepfakes are impacting these processes, we set out to investigate:

RQ2. To what extent do fact-checkers adapt their work routines to deepfakes in terms of (1) detection techniques, and (2) countermeasures?

28. [It's Nothing but a Deepfake! The Effects of Misinformation and Deepfake Labels Delegitimizing an Authentic Political Speech](#)

29. [Examining Misinformation and Deep Fakes](#)

30. [Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database](#)

Research Questions:

1. How actors or institutions are incentivized to promote misinformation/disinformation?

2. How misinformation is shared, promoted, and accessed through various channels?
3. How different types and characteristics of policy frames or narrative strategies can strategically shape views of issues, politicians, or social groups?
4. How individual and group political psychology, including in-group and out-group behavior, influence the efficacy of misinformation?
5. How the timing of messaging and counter-messaging leads to persistence effects or decay effects as contrasting messages compete or are reinforced?
6. How observational, quasi-experimental, and experimental evidence can be mustered to test the validity of theories, including studies of the efficacy of policy changes, mitigation efforts, and choices around AI system design?

31. [Political Deepfakes Are as Credible as Other Fake Media and \(Sometimes\) Real Media](#)

32. [Deep Fakes and Digital Ethics: Navigating the Ethical Dilemma in Portraying Political Characters through Deep Fakes and their Impact on Political Communication in Digital Media](#)

Research Objectives

- To examine various deepfake videos of popular Pakistani politicians on TikTok.
- To explore how people's trust in political leaders and institutions is changed by deepfakes.
- To study the problems that arise from this kind of content against popular politicians.

Research Questions

1. What kinds of fake content are developed by AI around political leaders on TikTok?
2. How much do deepfakes change people's beliefs about political leaders, the media, and various institutions?
3. What moral questions come up when deep-fake videos are viral against popular politicians?

Sampling Technique

We used purposive sampling to choose content that displayed a large and diverse variety of political statements. The reason for using this technique was that it makes it possible to focus on types of content that are directly connected to the research questions. Four main areas of focus shaped what the video sample contains:

1. Satire and Humor:

Videos that imitate political figures and sometimes overstate their personality or opinions.

2.Misinformation: Videos containing scenarios or remarks that didn't take place, designed to see how easily viewers can detect false news.

3.Motivating or Nationalistic Videos: Movies showing heads of nations giving inspiring speeches to measure their impact on viewers.

4.Edited Renderings of Real Speeches: Political messages are changed to include opinions that represent opposing or controversial thoughts.

They discussed the analysis of 40 deepfake videos created for the four selected Pakistani political figures in detail. The videos were sorted and tagged into groups:

- **Satirical Parodies,**
- **Fake Political Statements,**
- **Manipulated Campaigning,**
- **Inspirational Leadership,**
- **and Negative Propaganda.**

33. [Manipulated Minds: Political and Social Factors Enabling Deepfake Impact in Pakistan](#)

Research Objectives

1. To explore the social and political conditions that make Pakistani society vulnerable to deepfake influence.
2. To analyse how media consumption habits contribute to the acceptance and spread of manipulated digital content
3. To understand how social media echo chambers shape public susceptibility to deepfake-driven misinformation.

Research Questions

RQ1. How social and political conditions make Pakistani society vulnerable to deepfake influence

RQ2. How media consumption habits contribute to the acceptance and spread of manipulated digital content?

RQ3. How social media echo chambers shape public susceptibility to deepfake-driven misinformation.

33. [Deepfake Deception and the Need for Media Literacy in Pakistan.](#)

Research Objectives

RO1: To identify if, there is a difference in the level of understanding of new media-literates and non-new media-literates in distinguishing between deepfakes.

RO2: To examine the changes in trust levels regarding online videos in media-literate and non-media-literate people after exposure to deepfake videos.

Research Questions

RQ1: Do new media-literate and non-media-literate people differ in their capacity to identify deepfakes?

RQ2: To what extent does the witnessed exposure of deepfake videos affect the trustworthiness of media-literate and non-media-literate people in online videos?

Theoretical Framework

This study integrates Media Literacy Theory and Trust and Credibility Theory to understand the effects of NML on the capability of youngsters in Punjab, Pakistan to detect and address deepfakes. This framework helps examine how media literacy can help identify deepfakes and how this changes the blind trust in online content, fulfilling the purpose of the study, which focuses on how deepfakes impact perception and behaviour among the Pakistani youth.

34. [Psycho-social Impact of Deepfake Content in Entertainment Media](#)

Objectives :

- To Examine Emotional and Cognitive Responses
- To Explore Trust and Deception Dynamics
- To Evaluate Ethical Considerations
- To Assess Legal and Regulatory Frameworks
- To Analyze Practical Implications and Solutions

Research Questions

1. Does the deepfake technology significantly impact the psycho-social aspects of audience engagement with entertainment media?
2. What might be the key concerns regarding the psychological effects of deepfake content featuring well-known personalities?
3. How does the widespread acknowledgment of deepfake content as a threat to media credibility contribute to our understanding of the psycho-social effects associated with deepfake technology?

35. [Studying the Online Deepfake Community](#)

The common topic areas found on r/deepfakes during our analysis were the same as what was found on MDF, consisting of:

- General Discussion.**Discussions about the technology in general, including discussions on deepfake ethics, legality, censorship/community suppression, and alternative application areas
- Technical Resources and Assistance.**Discussions relating to errors encountered, making deepfakes, issues occurring with the creation tools themselves, and suggestions for improvements to those creation tools
- Creation Resource Sharing.**Users sharing datasets used to create deepfakes,sharing resources that can be used to generate data for a model, and sourcing high-quality videos for data ripping
- Content Sharing.**Users sharing their deepfakes, and users requesting someone re-post a previously deleted deepfake
- Content Requests.**Discussions surrounding who should be deepfaked, who target should be swapped onto for a high-quality output, and deepfake requests for specific targets
- Meta Discussion.**Discussions about the platform itself, rules, milestones, and announcements

36. [Merging AI Incidents Research with Political Misinformation Research: Introducing the Political Deepfakes Incidents Database](#)

Descriptors	URL, File, Media format, Summary
Social Media	Views, Likes, Comments, Shares
Source	Name, Entity type, Location
Sharer	Name, Occupation, No. of followers
Target(s)	Name, Entity type, Response (if any)
Verification	Presentation as real or fake, Visual markings of false content, Watermark/label, External sources
Context & Content	Text from post, Context from post and comments, Harm type depicted (e.g., violence or discrimination), Communication goal (e.g., satire or reputational harm)
Real-world Connections	Context (e.g., scandal, election, or war), Reported real-world harms (e.g., evidence of actual violence or political interference)
Politics & Policy	Policy sector, Framing, Narrative

Table 1: Overview of Variables in PDID

37. [DEEFAKE TECHNOLOGY AND ITS IMPACT: ETHICAL CONSIDERATIONS, SOCIETAL DISRUPTIONS, AND SECURITY THREATS IN AI-GENERATED MEDIA](#)

Research Questions :

1. How do deepfakes affect public perception, democracy, and trust in social institutions?
2. What types of security threats arise from deepfake-enabled cybercrime and Misinformation?
3. How effective are current legal and policy frameworks in addressing deepfake-related risks?

38. [Unmasking deepfakes: a multidisciplinary examination of social impacts and regulatory responses](#)

Table 2 Vulnerabilities and targeting mechanisms of populations affected by deepfake technology

Vulnerable group	How they can be targeted
Women targeted in pornography	Used in non-consensual deepfake pornography, leading to psychological trauma and social stigma (P11, P8, P3)
Political figures	Deepfakes can fabricate statements or actions to tarnish reputations or influence political outcomes. Furthermore, the spread of misinformation through deepfakes can lead to political instability, social unrest, or distrust in institutions, significantly disrupting the political and social spheres (P11, P7)
People lacking knowledge in AI	Vulnerable due to limited awareness and knowledge in identifying manipulated content, leading to the acceptance of misinformation and fraud (P4, P5)
Minorities and younger generations	Deepfakes exacerbate identity crises and suicidal attempts, with voice simulators and personal data used for fraud. This particularly threatens teenagers and young adults by fostering body dysmorphia and depression due to unrealistic online ideals (P1, P13)
Celebrities	Targeted for deepfake revenge porn and other manipulations, often disproportionately affecting women celebrities, leading to reputational damage and emotional distress (P11)
Businesses	Financial fraud and reputation damage through deepfake impersonations of executives or misleading information (P7)
The elderly	Elderly individuals are more susceptible to scams and manipulative tactics due to factors like trust, vulnerability, and less familiarity with emerging technologies. Their potential lack of awareness about deepfakes' nature can lead to direct and severe manipulation (P14, P5, P8)

39. [Deepfakes: The Legal Implications](#)

Deepfake technology has progressed at an alarming rate in the past decade and appear in three very general categories which includes:

Face Swapping

Face swapping involves or pertains to the replacing one's person's face with that of another in either a photo or video content.

Lip Syncing

Lip syncing pertains to the making of another person appear to say something they are not in audio or video content.

Puppet Technique

The Puppet technique refers to the fake movements of an individual in an unnatural manner

40 . [The Web of False Information: Rumors, Fake News, Hoaxes, Clickbait, and Various Other Shenanigans](#)

Types of False Information

False information on the Web can be found in various forms, hence we propose the categorization of false information into eight types as listed below:

- **Fabricated (F) [154]**. Completely fictional stories disconnected entirely from real facts. This type is not new and it exists

since the birth of journalism. Some popular examples include fabricated stories about politicians and aliens [84] (e.g., the story that Hillary Clinton adopted an alien baby).

- **Propaganda (P) [101]**. This is a special instance of the fabricated stories that aim to harm the interests of a particular party and usually has a political context. This kind of false news is not new, as it was widely used during World War II and the Cold War. Propaganda stories are profoundly utilized in political contexts to mislead people with the overarching goal of inflicting damage to a particular political party or nation-state. Due to this, propaganda is a consequential type of false information as it can change the course of human history (e.g., by changing the outcome of an election). Some recent examples of propaganda include stories about the Syria air strikes in 2018 or about specific movements like theBlackLivesMatter (see [124] for more examples).

- **Conspiracy Theories (CT) [69]**. Refer to stories that try to explain a situation or an event by invoking a conspiracy without proof. Usually, such stories are about illegal acts that are carried out by governments or powerful individuals. They also typically present unsourced information as fact or dispense entirely with an

“evidence” based approach, relying on leaps of faith instead. Popular recent examples of conspiracy theories include the Pizzagate theory (i.e., Clinton’s campaign running a pedophile ring) [188] and conspiracies around the murder of Seth Rich [187] (e.g., Seth Rich was involved in the DNC email leaks).

- **Hoaxes (H) [109]**. News stories that contain facts that are either false or inaccurate and are presented as legitimate facts. This category is also known in the research community either as half-truth [3] or factoid [14] stories. Popular examples of hoaxes are stories that report the false death of celebrities (e.g., the Adam Sadler death hoax [164]).

- **Biased or one-sided (B)**. Refers to stories that are extremely one-sided or biased. In the political context, this type is known as Hyperpartisan news [144] and are stories that are extremely biased towards a person/party/situation/event. Some examples include the widespread diffusion of false information to the alt-right community from small fringe Web communities like 4chan’s /pol/ board [86] and Gab, an alt-right echo chamber [199].

- **Rumors (R) [140]**. Refers to stories whose truthfulness is ambiguous or never confirmed. This kind of false information is widely propagated on OSNs, hence several studies have analyzed this type of false information. Some examples of rumors include stories around the 2013 Boston Marathon Bombings like the story that the suspects became citizens on 9/11 or that a Sandy Hook child was killed during the incident [163].

- **Clickbait (CL) [56]**. Refers to the deliberate use of misleading headlines and thumbnails of content on the Web. This type is not new as it appeared years before, during the “newspaper era,” a phenomenon known as yellow journalism [50]. However, with the proliferation of OSNs, this problem is rapidly growing, as many users add misleading descriptors to their content with the goal of increasing their traffic for profit or popularity [143]. This is one of the least severe types of false information because if a user reads/views the whole content then he can distinguish if the headline and/or the thumbnail was misleading.

- **Satire News (S) [48]**. Stories that contain a lot of irony and humor. This kind of news is getting considerable attention on the Web in the past few years. Some popular examples of sites that post satire news are TheOnion [22] and SatireWire [16]. Usually, these sites disclose their satyric nature in one of their pages (i.e., About page). However, as their articles are usually disseminated via social networks, this fact is obfuscated, overlooked, or ignored by users who often take them at face value with no additional verification.

41. Proposing a practical taxonomy of misinformation for intervention design

Table I: Types and Strategies of Misinformation

Advertising	Error	Hyperbole	Opinion masquerade as fact
Biased/one-sided	Fabrication	Imposter	Parody/Satire
Bogus	Factitious information blends	Junk news	Persuasive information
Cherry picking	Fallacy	Leaks	Propaganda
Clickbait	False balance	Lie	Pseudoscience
Conspiracy theories	False connection	Lying by omission	Quackery
Data manipulation	False context	Malinformation	Rumors
Deep fakes	False information	Manipulation	Sensationalism
Defamation	False/Fake news	Manufactured amplification	Systemic lies
Discredited research	Fearmongering	Misappropriation	Trolling
Disinformation	Forgeries	Misinformation	Typosquatting
Doxing	Hate speech	Misleading content	Urban legend
Echo chamber	Hoax	Misquote	

42. Deception and Manipulation in Generative AI

Deception and manipulation as misleadingness We can now characterize deceptive and manipulative behavior in LLMs (and other language-generating AI systems) as follows:

A statement¹⁰ is deceptive with respect to question Q if it tends to move its addressee's beliefs about Q further away from the beliefs that she would endorse under semiideal conditions.

A statement is manipulative with respect to a choice C if it tends to move its addressee's behavior in choice C further away from the behavior that she would endorse under semi-ideal conditions.

Tarsney, C. (2024). Deception and Manipulation in Generative AI. In *arXiv [cs.CY]*.

<http://arxiv.org/abs/2401.11335>

Proposed Questionnaire :

Fact-checkers framework

Encounter Deepfakes (Experience with AI generated Contents)

Responding to Deepfakes (Detection Tools and techniques)

Understanding Deepfakes (Verification and Anticipation of future trends)

Concern Category

- **Authenticity Concern**
- **Spread Concern**
- **Influence Concern**
- **Accountability Concern**

- a. [What are the key trends in the creation and dissemination of AI-generated misinformation including topics, origins, targets, intentions, and generative models?](#)
- b. [What detection methods and debunking strategies are employed by fact-checkers and what additional insights do they provide for countering AI-generated misinformation?](#)
- c. [How do fact-checkers assess the ongoing spread of AI-generated misinformation and how do they address the challenges it poses?](#)
- d. [What dangers do fact-checkers identify in connection with deepfakes for \(i\) Journalism \(ii\) audience?](#)
- e. [What extent do fact-checkers adapt their work routines to deepfakes in terms of \(i\) detection technique \(ii\) counter measures ?](#)
- f. [What are the primary identification tools currently accessible and employed by fact-checking organizations?](#)

- g. How does the widespread acknowledgment of deepfake content as a threat to media credibility contribute to our understanding of the psycho-social effects associated with deepfake technology?
- h. Which patterns emerge in the authorship and geographical focus of research on fact-checking tools in algorithm-driven journalism?
- i. To what extent to which Fact-checkers are prepared for deepfakes as a new form of disinformation?
- j. What kinds of fake content are developed by AI around political leaders on Facebook?
- k. How much do deepfakes change people's beliefs about political leaders, the media, and various institutions?
- l. Does the deepfake technology significantly impact the psycho-social aspects of audience engagement with entertainment media?
- m. To what extent does the witnessed exposure of deepfake videos affect the trustworthiness of media-literate and non-media-literate people in online videos?

Questions for Actors

- 1. How actors or institutions are incentivized to promote misinformation/disinformation?
- 2. How misinformation is shared, promoted, and accessed through various channels?
- 3. How individual and group political psychology, including in-group and out-group behavior, influence the efficacy of misinformation?

Creation and spread of AI-Generated Misinformation

1. What recent trends have you observed in the types of AI-generated misinformation you encounter?
2. Where do you usually see these contents originate from, and who do they seem to target the most?
3. Have you noticed specific themes (e.g., politics, health, religion) becoming more common in AI-generated misinformation?

Detection and Debunking Strategies

4. What tools or methods do you use to detect whether a content is AI-generated or a deepfake?
5. Can you describe the process you follow when fact-checking a suspected AI-generated content?
6. Are there any particular techniques that are more effective in debunking deepfakes or synthetic content?

Challenges and Response Tactics

7. What are the main challenges you face when trying to identify and counter AI-generated misinformation?
8. Do you think your organization is adequately equipped to handle this new wave of misinformation?

Deepfakes and Their Impact

9. In your experience, how have deepfakes impacted public trust in journalism or media outlets?
10. Do you believe that audiences are more vulnerable or skeptical toward deepfakes? How do they usually react?
11. Have you had to modify your own workflow or team structure to better respond to the threat of deepfakes?

Tools and Technologies

- 12. What kind of AI or detection tools are you currently using to identify deepfakes or synthetic content?
- 13. Are there any tools or features you wish existed that could make fact-checking deepfakes easier?

Preparedness and Institutional Readiness

- 14. To what extent do you feel fact-checkers as a community are ready to tackle deepfakes as a new form of disinformation?
- 15. Is there any collaboration between organizations or platforms to better respond to AI-generated misinformation?

Societal and Psychological Effects

- 16. Have you noticed any psychological or social effects in the public as a result of repeated exposure to deepfakes?
- 17. Do you think deepfakes are creating more political polarization or distrust in institutions?

Patterns in Misinformation Authorship

- 18. Are there specific regions, groups, or actors that appear more frequently in creating or spreading deepfake content?
- 19. Do you see a difference in how misinformation is generated or received in urban vs rural audiences or across countries?

Platform-Specific Observations

- 20. What kinds of AI-generated political content are you seeing specifically on platforms like Facebook?

21. Are there differences in how deepfakes are presented or received on Facebook versus other platforms like YouTube or TikTok?

Actors

I. Motivation and Intention

1. What motivated you to start creating deepfake content? Was it satire, political messaging, artistic expression, or something else?
(aligns with "Satirical Parody", "Fake Political Statements", "Negative Propaganda")
2. Do you consider your content a form of entertainment, activism, or manipulation? Why?
3. Have you ever been approached by political or commercial actors to produce tailored deepfakes?
4. Do you personally believe your content influences public opinion, and is that your intended goal?

II. Tools, Techniques, and Sampling Strategy

5. What tools or platforms (e.g., FaceSwap, DeepFaceLab, Runway, etc.) do you use to create deepfakes?
(links to "Sampling Technique" and "Generative Models")
6. How do you decide which real-life video or clip to manipulate? Is it based on trending topics or targeted figures?
(refers to "Platform", "Situation", "Interaction")
7. Which content style do you prefer—Face-swapping, Lip-syncing, Puppet motion, or Manipulated Reality—and why?
8. Do you design content for specific platforms (Facebook, YouTube Shorts, TikTok), or does your content evolve across platforms?

III. Audience and Impact Understanding

9. Who do you think your audience is—general public, youth, political supporters, or opposition groups?

10. Have you ever received messages or comments that revealed your audience believed your deepfake content was real? How did you respond?

11. Do you monitor how far your content spreads or track engagement metrics (views, shares, comments)?

12. Have any of your contents been flagged, fact-checked, or removed? How did you react?

IV. Ethics, Accountability, and Boundaries

13. How do you distinguish between satire and harmful misinformation in your content creation?

14. Do you set any personal or ethical boundaries while creating deepfakes—such as avoiding sensitive individuals, religious symbols, or false crime scenarios?

15. Do you think creators should be held accountable for the influence or spread of their content? Why or why not?

16. What's your view on the growing concern around the authenticity of media and public trust due to deepfakes?

V. Future Trends and Evolution

17. Do you think the tools and models for deepfakes are becoming more accessible and realistic? How do you see this changing your work?

18. What role do you think deepfake creators like yourself will play in the future of media and political storytelling?

19. Do you anticipate any shifts in regulation or social media platform policies that may limit or support your work?

20. Would you ever collaborate with researchers or platforms to improve transparency around synthetic content?

Findings:

This pilot study reveals the layered complexity of AI-generated misinformation. The posts analyzed show that misinformation is not limited to spreading falsehoods but also functions as a tool of **persuasion, influence, and manipulation**. Unlike traditional disinformation, AI-driven content operates on two levels simultaneously: **deception**, where audiences are led to believe in fabricated realities, and **manipulation**, where narratives are strategically framed to shape opinions, reinforce biases, or provoke emotional responses.

The framework's initial pilot highlighted these challenges, demonstrating that existing approaches to misinformation research are often insufficient for capturing the hybrid nature of AI content. By merging two frameworks and integrating new elements, we were able to build a more comprehensive model that acknowledges both **narrative intent** and **audience impact**.

Preliminary observations also suggest that posts tagged with fact-checking articles often gain less engagement compared to unchecked misinformation, underscoring a **visibility gap in corrections**. This indicates that while fact-checkers play an important role, their reach is limited compared to the virality of manipulative posts.

Overall, the study emphasizes the urgent need for updated frameworks and methodologies to address AI-driven misinformation. The refined model developed through this pilot now provides a stronger basis for analyzing how such content circulates, how it is received, and how fact-checking practices can adapt to meet these challenges.

Limitations and How I overcome them

Uncertainty in Categorization Framework

- **Limitation:** At the start of data collection, the framework for classifying AI-generated misinformation was not finalized, creating ambiguity in systematic organization.
- **Overcome:** We began with a provisional framework and refined it in parallel with a literature review. The final framework was grounded in established academic studies.

Simultaneous Data Collection and Literature Review

- **Limitation:** Data collection started before the framework was fully validated, which risked inconsistency in classification.
- **Overcome:** The framework was continuously updated as the literature review progressed, ensuring alignment with theoretically sound categories.

Need for Validation of Framework

- **Limitation:** The initial framework lacked formal testing, raising concerns about its reliability.
- **Overcome:** Two pilot tests were conducted. The first highlighted classification challenges, and the second incorporated improvements based on global studies. This confirmed the framework's effectiveness.

Dual Nature of AI-driven Misinformation

- **Limitation:** AI-generated misinformation often functions both as **deception** (creating false beliefs) and **manipulation** (influencing attitudes), making it harder to classify under traditional categories.
- **Overcome:** I integrated these two dimensions into the framework and merged them with an existing model of eight variables, adding extra elements such as engagement metrics and narrative framing.

Project 3

Project Name: Disinformation Targeting Female Political Figures

Start Date: 21 July 2025

End Date: 14 August 2025

Project Overview:

This project investigated disinformation campaigns targeting political figures in Bangladesh, with a particular focus on how manipulated content exploits gender, political identity, and social divisions. Using fact-checking reports and social media monitoring, the study examined claims ranging from fabricated and altered content to deepfakes, hate speech, and coordinated harassment. Each case was analyzed in terms of narrative framing, target profile, dissemination strategies, and potential harm (psychological, reputational, electoral, or incitement of violence). A special emphasis was placed on gendered disinformation, identifying how female politicians, activists, and journalists are disproportionately subjected to misogyny, sexualization, and harassment in digital spaces. By combining local fact-checking data with international literature on political disinformation and gendered misinformation, the project aimed to build a framework for understanding the complex ways political actors are attacked online.

My Tasks and Responsibilities:

In this project, I collected and cleaned fact-checking reports on disinformation targeting political figures and categorized them by type, including fabricated claims, altered images, deepfakes, and harassment. I identified targeted individuals, linked disinformation to specific events, and analyzed how content spread across platforms and echo-chambers. I also cross-checked entries with fact-checking verdicts from major organizations to highlight verification gaps and contributed to refining the framework by integrating dataset findings with relevant global literature.

Data Collection: Fact-checking reports and disinformation cases were collected and analyzed during the project period. Posts were examined across Facebook, Twitter, Threads, TikTok, and other platforms, focusing on misinformation targeting political figures.

Content Classification: Each case was categorized by **type of claim** (fabricated content, deepfakes, edited images, hate speech, violent threats, doxing, unwanted explicit content). Gender-based narratives such as **misogyny, stereotypes, sexualization, and harassment** were marked separately to understand patterns of gendered disinformation.

Target Identification: Female Political leaders, party representatives, former MPs, debut politicians, and celebrities involved in politics were mapped as primary targets. The analysis traced **why they were attacked**, linking disinformation to events such as elections, protests, or scandals.

Spread Analysis: The project assessed **who spread the content**, its level of virality, and whether it was amplified through **eco-chambers** and coordinated networks. This enabled mapping of organized vs. individual disinformation flows.

Fact-Checking Comparison: Posts were cross-referenced with fact-checking verdicts from Fact-Watch, Rumor Scanner, Boom Bangladesh, and Dismislab. This provided insights into **coverage gaps**, inconsistencies, and areas where false content went unchecked.

Harm Modality Mapping: Each case was examined for potential harms, categorized into **psychological threats, reputational damage, incitement of offline violence, and electoral impact**.

Literature Integration: Global studies on **political astroturfing, scandals, cybersecurity frameworks, and gendered misinformation** were reviewed to align the framework with international scholarship.

Framework Refinement: Insights from the dataset and literature were merged to refine a **composite framework** capturing both political and gendered disinformation dynamics in Bangladesh.

Literature Review

1. Deepfakes and Harm to Women

Abstract : As deepfake technologies become more sophisticated and accessible to the broader online community, their use puts women participating in digital spaces at increased risk of experiencing violence online and abuse. In a ‘post-truth’ era, the ability to discern what is real and what is fake allows malevolent actors to manipulate public opinion or ruin the social reputation of individuals to wider audiences. While the scholarly research on the topic is sparse, this study explored the harm women have experienced in technology and deepfakes. Results of the study suggest that deepfakes are a relatively new method to deploy gender-based violence and erode women’s autonomy in their on-and-offline world. This study highlights the unique harms for women that are felt on both an individual and systemic level and the necessity for further inquiry into online harm through deepfakes and victims’ experiences.

Source : https://www.researchgate.net/publication/371862022_Deepfakes_and_Harm_to_Women

2. The emergence of AI-generated deepfakes as a new tool for gender-based violence against women: A brief narrative review of evidence and the implications of the techno-feminist perspective

Abstract : Technology facilitates crime. Perpetrators use technology to commit new crimes or old crimes in new ways. The advancement of technology has spawned several cybersexual crimes that affect women disproportionately. The emergence of deepfakes has brought a new dimension to women's cyber sexual victimization. There is little empirical research on deepfakes and women's sexual abuse. Therefore, for this comment piece we reviewed existing peer-reviewed articles and data from websites to show that deepfakes have become a new tool for gender-based violence against women. We use a techno-feminist perspective to argue that deepfake technology is an addition to the tools that are used to form a patriarchal-capitalist society. Like other cybercrimes, cybercrimes using deepfakes affect women disproportionately, and deepfakes have become a tool to control and subjugate women. Also, we contend that techno-feminism has ignored the masculine aspects of new technologies. There is a need for more research on the role of deepfake technology in women's sexual abuse to prevent gender-based violence against women.

Source : https://www.researchgate.net/publication/390743001_The_emergence_of_AI-generated_deepfakes_as_a_new_tool_for_gender-based_violence_against_women_A_brief_narrative_review_of_evidence_and_the_implications_of_the_techno-feminist_perspective

3. Is AI-Facilitated Gender-Based Violence the Next Pandemic? (Article)

Source : <https://www.theregreview.org/2024/05/06/de-silva-de-alwis-vialle-is-ai-facilitated-gender-based-violence-the-next-pandemic/>

4. Adverse human rights impacts of dissemination of nonconsensual sexual deepfakes in the framework of European Convention on Human Rights: A victim-centered perspective

Abstract : Generative artificial intelligence systems have advanced significantly over the past decade and can now generate synthetic but highly realistic audio, photo, and video, commonly referred to as deepfake. Image-based sexual abuse was the first widespread (mis)use of deepfake technology and continues to be the most common form of its misuse. However, further (empirical) research is needed to examine this phenomenon's adverse human rights implications. This paper analyses the potential adverse human rights impacts of the dissemination of nonconsensual sexual deepfakes in the framework of the European Convention on Human Rights and argues that the dissemination of such deepfakes can hinder the rights protected by the Convention. These include the right to respect for private and family life, as nonconsensual sexual deepfakes can undermine data protection, harm one's image and reputation, and compromise psychological integrity and personal autonomy. Additionally, such deepfakes can threaten freedom of expression by creating a silencing effect on public watchdogs, politicians, and private individuals. Finally, nonconsensual sexual deepfakes can impair the economic and moral rights of pornography performers by abusing their work and bodies to abuse others without authorization and compensation. These findings highlight that the Council of Europe member states must fulfil their obligations to provide effective protection against this technology-facilitated, gender-based, and sexual violence.

Source : <https://www.sciencedirect.com/science/article/abs/pii/S0267364925000032>

5. It's Everyone's Problem: Mainstreaming Responses to Technology-Facilitated Gender-Based Violence

Abstract : This report assesses the state of research on TFGBV as well as recent global legislative, regulatory, and policy progress made on this issue. Through a case study that explores the data around the online harassment and abuse of Australian eSafety Commissioner Julie Inman Grant, the report documents the real-life effects of TFGBV for women in public life. We argue that TFGBV is not an intractable problem, but one that must be mainstreamed to be mitigated, centering women's experiences in broader policy debates. TFGBV must no longer be the responsibility solely of women's advocacy groups. Technology companies, governments, civic tech organizations, law enforcement, employers, schools, and others must mainstream their work to combat TFGBV to reflect its mainstreamed effects on society. To this end, we recommend a number of practical solutions to the specific and pressing issues that women and girls face online today. Addressing the urgent changes described here will not only make women and girls safer and ensure their voices are heard, but also improve the safety and free expression for everyone who uses the internet, building more robust, representative democracies.

Source : <https://apo.org.au/node/328240>

6. Technology-Facilitated Gender-Based Violence An Emerging Issue in Women, Peace and Security

Abstract : As the uses and abuses of technology develop at a rapid rate, the growing threat and consequences of technology-facilitated gender-based violence (TFGBV) must be addressed by those working to advance the Women, Peace and Security (WPS) Agenda. Existing policies and programs must be adapted and new ones adopted to account for digital dimensions of gender-based violence (GBV) and to keep online spaces safe for women.

Technology has helped advance women's rights around the world in important ways. Technological tools make education and employment more accessible, and online spaces create new opportunities to build movements and amplify calls for equality and justice. These gains serve the goals of the WPS Agenda, which aims to protect and empower women in conflict-affected contexts. Along with these benefits, however, are substantial costs. Globally, 66 percent of women

have reported experiencing TFGBV—including instances of cyber-harassment and stalking, doxxing, and image-based sexual abuse.¹ Further, there are increasing reports of digital tools being used to exacerbate offline violence. For example, sexual assaults of Iranian protesters have been filmed and used to blackmail the protesters,² and women in Pakistan have been killed as a result of photoshopped images being posted online.³ National governments and international organizations increasingly recognize TFGBV as a pressing issue. Existing research and policies focus on the impact of TFGBV on specific women—primarily, politicians and journalists—with less attention paid to the implications of TFGBV for the WPS Agenda, particularly with regard to protecting women from GBV and promoting their participation in peace processes. Given the scale, scope, and impact of TFGBV, the WPS field must evolve to account for the unique challenges women are facing in an increasingly digitized world. Based on a comprehensive desk review of existing TFGBV literature, this report explores the linkages between TFGBV and the goals of the WPS Agenda. The first section introduces the WPS Agenda, specifically focusing on its structure and the context for its creation. The next section links TFGBV with the protection goals of WPS, outlining the prevalence of existing types of digital violence and the challenges with addressing them. It further highlights the link between digital and offline violence, demonstrating that protection efforts must account for a continuum of violence. The third section draws out threats TFGBV poses to the participation goals of WPS. It identifies three key dimensions of TFGBV that hinder women's participation in peace and politics: targeted attacks against prominent women, the “gendered chilling effect” of online violence that discourages women from participating in online spaces, and online radicalization and misogyny.

7. NEW FORMS OF GENDER-BASED VIOLENCE: BETWEEN DEEPPREAL AND DEEPPFAKE

Abstract : This article constitutes an exploratory contribution to the understanding of the intersection between digital technologies and gender-based violence, with a particular focus on the emerging threats posed by *deepfake* technologies. It combines theoretical insights and empirical analysis based on data collected through a non-probabilistic opinion survey. The fundamental objective of this article is to analyze how *deepfake* and *deepreal* phenomena are perceived in the context of gender-based violence and to identify patterns in public awareness and attitudes toward these technologies. The methodological approach involves qualitative analysis supported by survey data, aimed at capturing trends rather than making generalizable conclusions. Through this approach, the article seeks to offer a critical discussion on the ethical, social, and cultural ramifications of *deepfake*-related gender violence, paving the way for future research and regulatory recommendations.

Source : <https://studiahumanitatisjournal.com/revista/index.php/shj/article/view/110/195>

8. A feminist legal analysis of non-consensual sexualized deepfakes: contextualizing its impact as AI-generated image-based violence under EU law

Abstract : This article focuses on the legal contextualization of the generation of fake audiovisual sexual content based on replacing and shuffling facial or voice data using artificial intelligence (AI) technology without the previous consent, commonly known as ‘non-consensual sexualized deepfakes’. The legal analysis will be limited to examination of the impact of non-consensual sexualized deepfakes on the gender equality spectrum and gender-based violence context. For this reason, the analysis will be built upon the European AI Act – the applied legal framework when AI software is used – the European Directive on combating violence against women and domestic violence (GBV Directive) – which criminalizes the generation of sexualized deepfakes – and the Digital Service Act (DSA) – the legislative framework regulating online platforms, where

mostly sexualized deepfakes are distributed. To better understand how sexualized deepfakes are generated, we will examine their generation and detection techniques through a short overview from a technical lens. The methodology used is built on the feminist critical approach – based on the ‘King Kong’ theory and abolitionist feminism – to the doctrinal analysis of the relevant European legislation. After having captured the harms of deepfake technology, legal solutions will be provided through this cross-disciplinary perspective.

Source : <https://www.tandfonline.com/doi/abs/10.1080/23268743.2024.2408277>

9. Are Deepfakes Concerning? Analyzing Conversations of Deepfakes on Reddit and Exploring Societal Implications

Abstract : Deepfakes are synthetic content generated using advanced deep learning and AI technologies. The advancement of technology has created opportunities for anyone to create and share deepfakes much easier. This may lead to societal concerns based on how communities engage with it. However, there is limited research available to understand how communities perceive deepfakes. We examined deepfake conversations on Reddit from 2018 to 2021—including major topics and their temporal changes as well as implications of these conversations. Using a mixed-method approach—topic modeling and qualitative coding, we found 6,638 posts and 86,425 comments discussing concerns of the believable nature of deepfakes and how platforms moderate them. We also found Reddit conversations to be pro-deepfake and building a community that supports creating and sharing deepfake artifacts and building a marketplace regardless of the consequences. Possible implications derived from qualitative codes indicate that deepfake conversations raise societal concerns. We propose that there are implications for Human Computer Interaction (HCI) to mitigate the harm created from deepfakes.

Source : <https://arxiv.org/pdf/2203.15044>

Dataset : [Disinformation Targeting Political Figures](#)

After Cleaning Analysis_Disinformation Targeting Political Figures .xlsx									
File Edit View Insert Format Data Tools Help									
A18									
	A	B	C	D	E	F	G	H	I
17	Area			Meta Data		Verdict & Evidence	Target Characteristics		
18							Name		
19		31 Oct 2024	Rumor Scanner	https://rumorsscanner.com/fact-check/maek-hamans-arrested-false-claim-127379	ফারহান নেভী আফিকাকে কারকে বিন্দু খবরকে প্রচার	False	Atika Binte Hossain	Student Leader	Former Student League Leader
20		26 Aug 2024	Rumor Scanner	https://rumorsscanner.com/fact-check/maek-hamans-arrested-false-claim-127379	ফেরি হাফিজের ফটোকার্ড নকল করে ফারহান নেভী আফিকাকে নিয়ে ভুয়া তথ্য	False	Atika Binte Hossain	Student Leader	Student league leader
21		31 Aug 2024	Rumor Scanner	https://rumorsscanner.com/fact-check/maek-hamans-arrested-false-claim-127379	মহারাজের সংবাদবিত্তার খবরকে আফিকাকে জরিদে মদুন টিকির ভুয়া ফটোকার্ড প্রচার	False	Atika Binte Hossain	Student Leader	Student league leader
22		2 June 2024	Fact-Watch	https://www.fact-watch.org/take-snapshot	সাবেক ফারহান নেভীর নামে খব নিয়ে খবর প্রচার ভুয়া তথ্য প্রচার	Altered	Atika Binte Hossain	Student Leader	Student League Leader
23		24 Nov 2024	Rumor Scanner	https://rumorsscanner.com/fact-check/maek-hamans-arrested-false-claim-127379	আফিক নজরুলের সাথে অভিনেত্রী বীথনের এই ছবিটি সম্পর্কিত	Altered	Azmi Haque Badhon		Actress
24		16 Nov 2024	Rumor Scanner	https://rumorsscanner.com/fact-check/maek-hamans-arrested-false-claim-127379	অফিক নেভী বীথন উপন্যাসী হাফিজ খবরকে জরিদে মদুন টিকির নামে ভুয়া ফটোকার্ড প্রচার	False	Azmi Haque Badhon		Actress

Findings: This project revealed that disinformation targeting political figures in Bangladesh is both strategic and gendered. Manipulated content—ranging from altered images and videos to fabricated claims and hate speech—was frequently used to damage reputations, reinforce partisan divides, and provoke strong emotional reactions among audiences. The analysis showed that gender-based attacks were especially prominent, with female politicians, activists, and journalists disproportionately targeted through misogyny, sexualization, and harassment. Such patterns contribute to a culture of digitally mediated political violence and raise concerns about the chilling effect on women’s political participation.

The spread analysis highlighted that disinformation is rarely random; instead, it often circulates through coordinated networks and echo-chambers, suggesting organized attempts to amplify harmful narratives. These campaigns were closely tied to political events, scandals, and electoral moments, showing how disinformation operates as a tactical weapon in political discourse.

Despite the seriousness of the content, the project found that fact-checking coverage was limited. Many viral and emotionally charged posts remained unverified, underscoring gaps in the current fact-checking ecosystem. The findings also mapped the harm modalities of such campaigns, including psychological threats, reputational damage, incitement of offline violence, and electoral consequences.

By combining these insights with global literature, the project identified an emerging typology of political disinformation in Bangladesh, one that blends political manipulation with gendered harassment and coordinated amplification. This typology provides a stronger foundation for developing frameworks that can track, categorize, and counter disinformation more effectively in both local and global contexts. **Overlapping Categories of Gendered Claims**

Limitations and How I overcome them

- **Limitation:** Many cases involved multiple forms of gender-based disinformation (e.g., misogyny combined with sexualization or harassment), making it difficult to assign them to a single category.
- **Overcome:** I created a coding system that allowed multiple labels for one case. This ensured nuance in classification without losing precision.

Verification Gaps in Fact-Checking

- **Limitation:** A large portion of viral disinformation cases were not covered by existing fact-checking organizations, limiting opportunities for cross-validation.
- **Overcome:** Instead of excluding these items, I treated them as **critical insights** showing verification gaps. This enriched the findings by identifying weaknesses in the fact-checking ecosystem.

Ambiguity in Political Target Profiling

- **Limitation:** Some actors had overlapping identities (e.g., a celebrity also involved in politics), creating uncertainty in classification.
- **Overcome:** I developed a hierarchy of priority (political role > activist identity > celebrity status) to resolve ambiguities systematically.

Subjectivity in Harm Assessment

- **Limitation:** Assessing harm modalities (e.g., reputational vs. psychological harm) involved subjective judgments.
- **Overcome:** I triangulated assessments with literature on political disinformation harms and cross-reviewed decisions with other team members for consistency.

Project 4

Project name: Mapping Actors Interfering in Online Discourse

Start Date: 24 July, 2025

End Date: On going

Project Overview:

This project focused on identifying and categorizing online actors who interfered with digital discourse following the events of **5 August 2025**. These actors were found to be spreading **misinformation, influence operations, and gender-based hate speech** on social media platforms, primarily Facebook. The dataset included pages, groups, and individual profiles that engaged in partisan narratives, extremist rhetoric, and targeted campaigns against political figures and activists. Using the **DMET framework (Disinformation–Misinformation–Extremism Typology)**, each actor was classified into levels ranging from **partisan accounts** to **fringe groups** and **violent extremist networks**. Additionally, the study documented how gendered narratives were systematically deployed to harass female activists, journalists, and political leaders, showing that online disinformation in Bangladesh is not only political but also deeply gendered. By compiling, filtering, and categorizing these actors, the project built a reference framework for understanding how digital manipulation networks operate in post-crisis political contexts.

My Tasks and Responsibilities:

In this project, I was responsible for collecting and organizing actor data, which included Facebook accounts, pages, and groups engaged in spreading disinformation. I applied the DMET classification system to categorize actors by their level of activity (partisan, fringe, extremist) and their affiliations. I also reviewed and filtered entries to identify key narratives, such as gender-based hate speech, violent extremism, and influence operations. Additionally, I cross-referenced the actor list with content samples to link specific posts to broader campaigns. My role further included assisting in building a clean and structured dataset by eliminating duplications, ensuring consistency, and refining categorizations for reliability. Through this work, I contributed to establishing a clearer map of online disinformation ecosystems and their role in shaping post-August 5th discourse.

Data Collection: Compiled a list of Facebook accounts, pages, and groups active after **5 August 2025**, focusing on those spreading misinformation, political propaganda, and gender-based hate speech.

Actor Classification: Applied the **DMET framework (Disinformation–Misinformation–Extremism Typology)** to categorize actors into levels (partisanship, fringe groups, violent extremism).

Content Analysis: Reviewed posts to identify **narratives, ideological leanings, and gender-based hate speech**, with emphasis on misogynistic and right-wing discourse.

Location Assessment: Attempted to trace the geographical origin of actors. Where no location was available, inference was made using post timing, language, and external reporting.

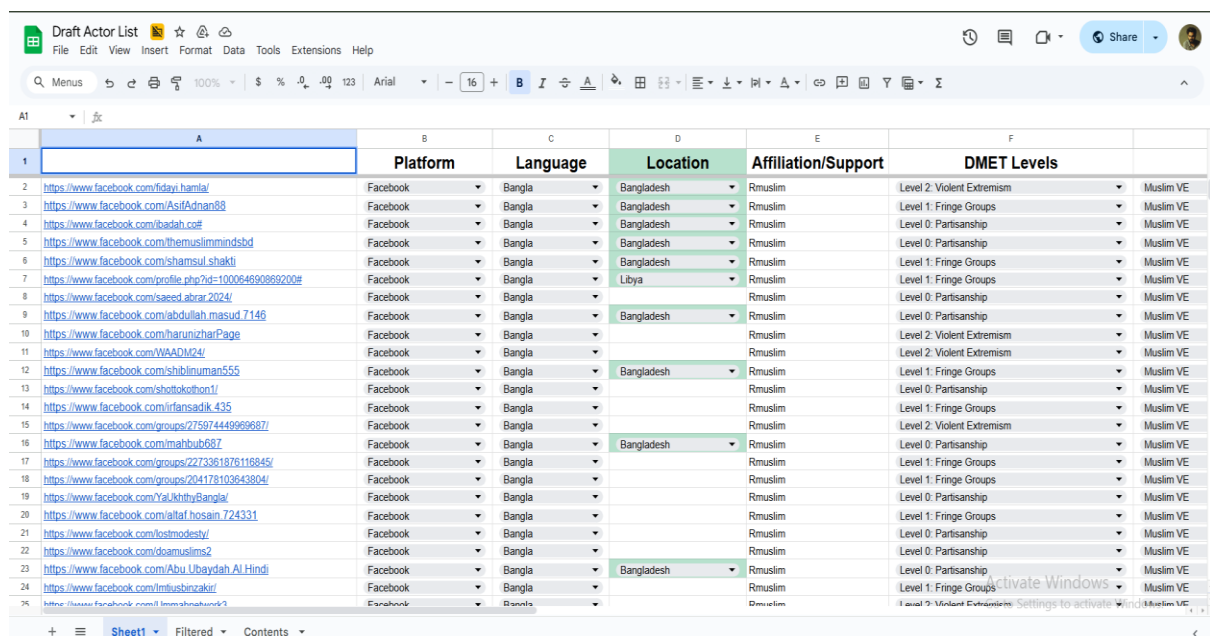
Identity Verification: For accounts with hidden or vague profiles, partisan leanings were determined by analyzing **post content, language, and shared networks**.

Cross-Referencing: Linked profiles with **known political affiliations**, particularly those tied to the former government and exiled MPs/ministers operating abroad.

Data Cleaning: Removed duplicate or inactive profiles and standardized entries across datasets to ensure reliability and consistency.

Framework Refinement: Adjusted categorization to allow for **multi-label narratives** (e.g., combining political propaganda with religious or nationalist framing) to capture complexity.

Dataset : [Mapping Actors Interfering in Online Discourse](#)



	A	B	C	D	E	F
		Platform	Language	Location	Affiliation/Support	DMET Levels
2	https://www.facebook.com/fidayi.hamla/	Facebook	Bangla	Bangladesh	Rmuslim	Level 2: Violent Extremism
3	https://www.facebook.com/AsifAdnan88	Facebook	Bangla	Bangladesh	Rmuslim	Level 1: Fringe Groups
4	https://www.facebook.com/badah.cof	Facebook	Bangla	Bangladesh	Rmuslim	Level 0: Partisanship
5	https://www.facebook.com/themuslimmindsbd	Facebook	Bangla	Bangladesh	Rmuslim	Level 0: Partisanship
6	https://www.facebook.com/shamsul.shakti	Facebook	Bangla	Bangladesh	Rmuslim	Level 1: Fringe Groups
7	https://www.facebook.com/profile.php?id=100064690869200#	Facebook	Bangla	Libya	Rmuslim	Level 1: Fringe Groups
8	https://www.facebook.com/saeed.abrar.2024/	Facebook	Bangla		Rmuslim	Level 0: Partisanship
9	https://www.facebook.com/abdullah.masud.7146	Facebook	Bangla	Bangladesh	Rmuslim	Level 0: Partisanship
10	https://www.facebook.com/harunizharPage	Facebook	Bangla		Rmuslim	Level 2: Violent Extremism
11	https://www.facebook.com/WAADM24/	Facebook	Bangla		Rmuslim	Level 2: Violent Extremism
12	https://www.facebook.com/shibinuman555	Facebook	Bangla	Bangladesh	Rmuslim	Level 1: Fringe Groups
13	https://www.facebook.com/shotokothon1/	Facebook	Bangla		Rmuslim	Level 0: Partisanship
14	https://www.facebook.com/irfansadik.435	Facebook	Bangla		Rmuslim	Level 1: Fringe Groups
15	https://www.facebook.com/groups/275974449969687/	Facebook	Bangla		Rmuslim	Level 2: Violent Extremism
16	https://www.facebook.com/groups/2273361876118845/	Facebook	Bangla	Bangladesh	Rmuslim	Level 0: Partisanship
17	https://www.facebook.com/groups/204178103643804/	Facebook	Bangla		Rmuslim	Level 1: Fringe Groups
18	https://www.facebook.com/YalukhtyBangla/	Facebook	Bangla		Rmuslim	Level 1: Fringe Groups
19	https://www.facebook.com/YalukhtyBangla/	Facebook	Bangla		Rmuslim	Level 0: Partisanship
20	https://www.facebook.com/altatf.hosain.724331	Facebook	Bangla		Rmuslim	Level 1: Fringe Groups
21	https://www.facebook.com/fostrmodesty/	Facebook	Bangla		Rmuslim	Level 0: Partisanship
22	https://www.facebook.com/doanmuslims2	Facebook	Bangla		Rmuslim	Level 0: Partisanship
23	https://www.facebook.com/Abu.Ubaydah.Al.Hindi	Facebook	Bangla	Bangladesh	Rmuslim	Level 0: Partisanship
24	https://www.facebook.com/imtushinazkir/	Facebook	Bangla		Rmuslim	Level 1: Fringe Groups
25	https://www.facebook.com/Imamshinazkir/	Facebook	Bangla		Rmuslim	Level 2: Violent Extremism

Findings

The analysis of actors interfering in online discourse after August 5 revealed several key patterns about the sources and nature of disinformation campaigns. The majority of accounts identified as spreading misinformation and manipulative narratives were **linked to individuals and networks affiliated with the former government**. These actors were consistently active in shaping online discussions, particularly around political legitimacy, national stability, and post-resignation narratives. Their coordinated presence indicates that pro-former government groups have maintained significant influence in Bangladesh's digital sphere, despite their loss of formal power.

A second important finding was the **role of gender-based hate speech**, which emerged predominantly from **right-wing networks**. These groups systematically targeted female activists, journalists, and political leaders with misogynistic slurs, sexualized attacks, and harassment campaigns. The prevalence of gendered narratives within political discourse highlights how gender is weaponized to discredit women in leadership and silence voices of dissent. Such campaigns not only perpetuate online abuse but also reinforce broader patterns of exclusion and intimidation against women in public life.

The dataset also revealed that several **former Members of Parliament and ministers aligned with the ex-government, who fled the country, remain active online**. Operating from abroad, they have used Facebook pages, groups, and other platforms to disseminate partisan content and mobilize narratives favorable to their political interests. Their continued digital presence illustrates the **transnational character of online disinformation**, where actors outside Bangladesh influence domestic discourse by leveraging social media networks.

Overall, the findings show that online interference after August 5 is not random but **politically coordinated**, with clear alignments along partisan and ideological lines. The dominance of pro-former government accounts, the gendered nature of right-wing attacks, and the involvement of exiled political elites demonstrate how online platforms have become contested spaces where political struggles are extended through misinformation and targeted harassment. These insights underscore the urgent need to monitor not only misinformation flows but also the **actors behind them**, as their strategies reveal both local and transnational dimensions of digital manipulation.

Missing Location Information

- **Limitation:** Some of the identified actors had no location data in their profiles, making it difficult to determine whether they were operating from inside Bangladesh or abroad. This created uncertainty in assessing the **transnational scope** of the disinformation campaigns.
- **Overcome:** I cross-referenced their activity patterns with external sources, such as news reports, page affiliations, and time-zone clues, to infer their likely location. In several cases, secondary reports confirmed that they were operating from outside the country, which allowed us to categorize them more accurately.

Hidden or Ambiguous Identities

- **Limitation:** A number of accounts did not openly reveal the identity of the individual behind them. This made it challenging to classify the actor's role and affiliations, especially when names or profile details were vague.
- **Overcome:** Instead of relying solely on profile data, I analyzed the **content of their posts**, including the language, narratives, and political alignment expressed. In most cases, the ideology and partisan leaning were clear from the discourse, which enabled accurate categorization despite hidden identities.

Overlap of Ideological Narratives

- **Limitation:** Some posts combined multiple narratives—such as anti-government rhetoric with religious or nationalist framing—making it difficult to assign them to a single category.
- **Overcome:** I applied a **multi-label categorization approach**, tagging these cases with more than one narrative type. This preserved analytical nuance and avoided oversimplifying complex ideological content.

Uncertainty in Attribution of Abroad-Based Actors

- **Limitation:** For actors presumed to be ex-government politicians or affiliates operating from abroad, direct attribution was difficult because official confirmations were often unavailable.
- **Overcome:** I reviewed their **posting networks, stylistic similarities, and cross-shared content** with verified partisan pages. By triangulating this evidence, I strengthened attribution, even where location data was absent.

Project 5

Project Name: Participation in BUP MediaComm Summit 2025

Date: 30 June – 01 July 2025

Venue: Bangladesh University of Professionals (BUP), Dhaka

Project Overview:

As part of my internship, I attended the BUP MediaComm Summit 2025, a two-day national convention hosted by the BUP Media and Communication Club and presented by Walton. The summit brought together students, educators, media professionals, and digital creators to explore the intersection of media innovation and communication practices. The event featured practical workshops, interactive sessions, competitions, and panel discussions. Key highlights included a workshop on digital content creation led by creators RK Shohan and Nasir Tamzid, and a session on fact-checking in the AI era that addressed challenges in combating misinformation.

The second day emphasized communication practices across sectors, with speakers such as filmmaker Redoan Rony, development communication specialist Nishath Sultana Purabi, and corporate representatives, who discussed strategies in media, NGO communication, and corporate communication. Alongside these sessions, competitions like Visual Storytelling and PitchCraft provided students with opportunities to showcase creative and strategic communication skills. The summit concluded with a prize-giving ceremony and closing remarks celebrating innovation and collaboration in communication.

My Tasks and Responsibilities

I attended the summit along with my supervisor, where I assisted him in designing and preparing presentation slides for his participation in the sessions. During the event, I attended the workshops and panel discussions, which provided me with valuable insights into the evolving role of digital content, fact-checking in the AI era, and the integration of communication strategies across different professional fields. This hands-on exposure helped me connect my academic and internship experiences with real-world practices, while also allowing me to observe how large-scale conventions foster learning and collaboration between students and professionals. Overall, the summit was an enriching experience that deepened my understanding of media and co

Communication in Bangladesh's contemporary landscape.



Learning Outcomes

Attending the **BUP MediaComm Summit 2025** was an enriching experience that expanded my perspective on both media practices and communication strategies. One of the key learnings was from the “**Legit Adda**” **Star Talk** featuring Ayman Sadiq, which emphasized the importance of **personal branding** and how social media is reshaping education and youth engagement. This session helped me understand how digital platforms are no longer just communication channels, but also spaces for **identity-building and influence**.

The workshops on **content creation** and **fact-checking in the digital age** provided practical skills that directly complemented my internship work. Learning about ideation, scripting, and basic video production from **Nasir Tamjid** and **RK Sohan** gave me insights into how storytelling can be translated into compelling digital formats. Equally, the fact-checking workshop conducted by **Shuvashish Dip** reinforced the urgency of combating misinformation and deepened my knowledge of digital verification methods—an area closely tied to my own research involvement at CQS.

The **panel discussion on media and communication in Bangladesh**, featuring professionals like **Redoan Rony, Zahedul Islam, and Nishath Sultana Purabi**, was particularly impactful. It highlighted how communication strategies differ across sectors—media, corporate, and NGOs—while also revealing common challenges such as adapting to technological change and meeting evolving audience expectations. This broadened my understanding of how communication operates in **multi-sectoral contexts**.

Finally, the closing session, “**Inside The Front Page**” by Fasbeer Eskander, sharpened my awareness of how **algorithms influence news curation** and the importance of **media literacy** for younger audiences. This reinforced the critical need to engage youth in conversations about how information is consumed and trusted in the digital era.

Overall, the summit enhanced my skills in **content creation, fact-checking, and communication analysis** while giving me the opportunity to interact with leading professionals and observe first-hand how theory and practice intersect in Bangladesh’s contemporary media landscape. It was a valuable learning platform that strengthened both my academic understanding and practical exposure.

Project 6

Project Name : DIU Students' Visit to FactWatch Office

Date : 16th July 2025

Project Description :

As part of my internship at FactWatch, I participated in hosting a visit by students from the **News Editing and Fact Checking course** at **Daffodil International University (DIU)**. The session aimed to introduce students to the practical processes of fact-checking, digital verification, and the organizational role of FactWatch in combating misinformation in Bangladesh. The event combined a presentation on FactWatch's mission, objectives, and methodologies with a **hands-on fact-checking workshop**, allowing students to learn directly from real examples of misinformation circulating online.

My Tasks and Responsibilities :

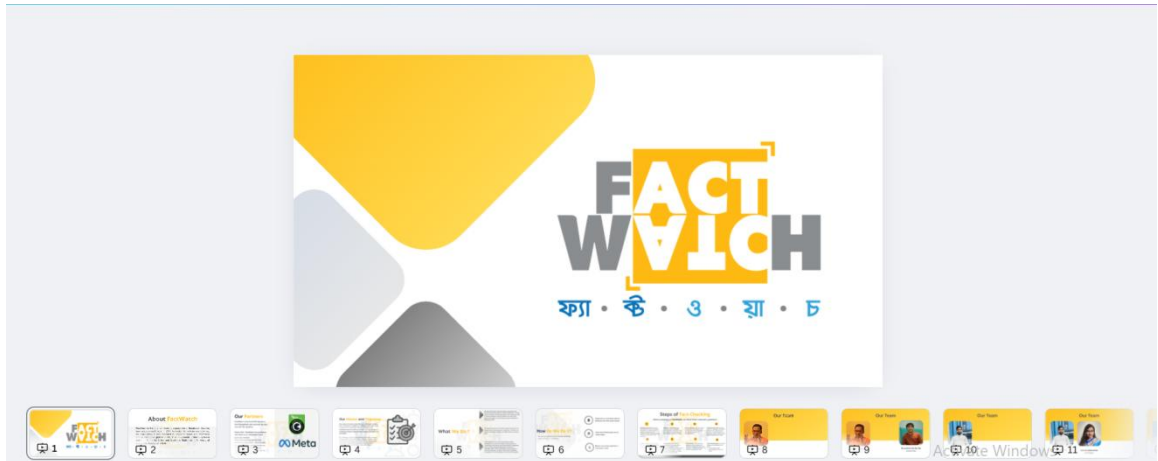
For this event, I was responsible for **designing the presentation slides** on behalf of the team, drawing on the provided company profile materials. I structured the slides to clearly highlight FactWatch's background, fact-checking process, and ongoing projects, ensuring that they were visually engaging and accessible to the students. During the session, I also **assisted in the fact-checking workshop**, supporting students in navigating digital tools, verifying claims, and understanding the principles of evidence-based journalism. This allowed me to apply my own training while contributing to the educational goals of the session.







Presentation Slides : [Fact-Watch Introduction](#)



Learning Outcomes

Participating in the DIU students' visit to FactWatch was an important learning experience that strengthened both my practical and interpersonal skills. By preparing and designing the presentation slides, I improved my ability to **translate organizational information into clear, structured, and visually engaging formats**. This gave me valuable experience in tailoring content for an academic audience that was new to fact-checking.

Assisting in the hands-on workshop allowed me to reinforce my own knowledge of **fact-checking tools and verification methods**, while also learning how to **guide others through the process step by step**. Explaining fact-checking techniques to students not only deepened my own understanding but also developed my skills in **mentorship and collaborative learning**.

Additionally, observing the curiosity and engagement of students gave me insight into how future journalists approach misinformation and the challenges they face in distinguishing credible sources. This experience highlighted the importance of **media literacy training** and the role of institutions like FactWatch in bridging the gap between theory and practice.

Overall, the session helped me grow as both a learner and a facilitator, combining technical expertise with communication skills in an educational setting.

Project 7

Project Name : ForensIQ Workshop – Digital Forensics 101: An Introduction to Digital Forensics & Open-Source Investigation in Bangladesh

Date : 31 July 2025

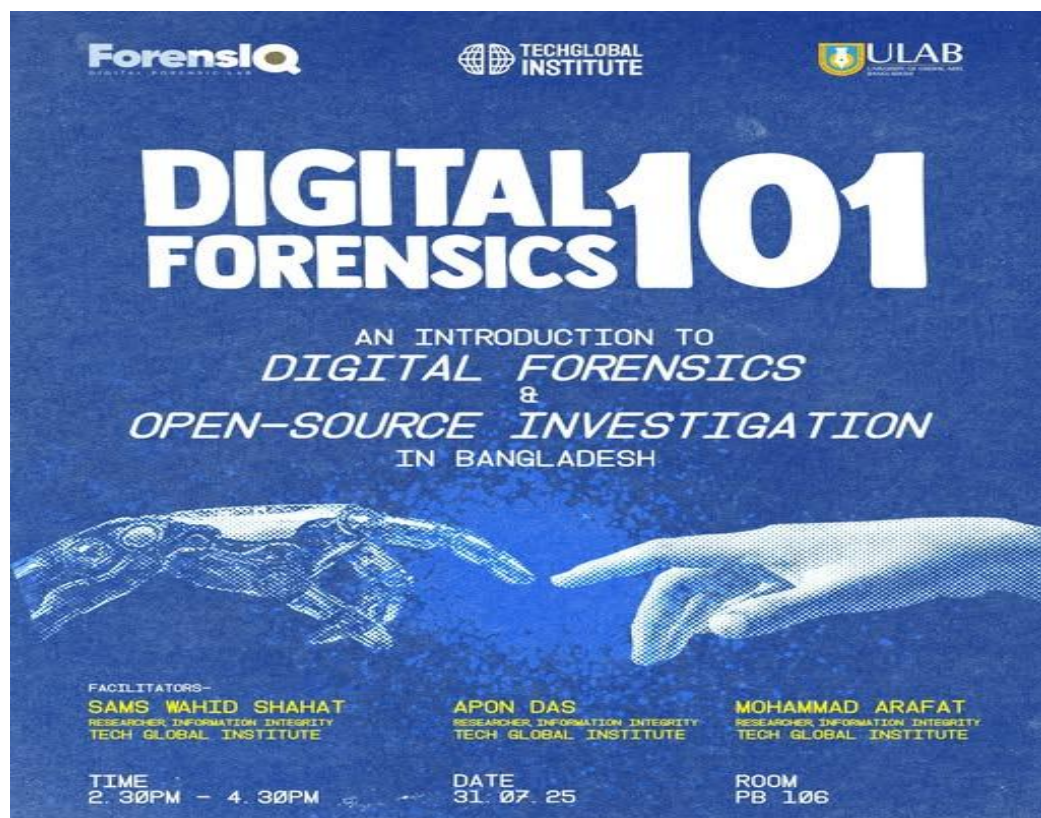
Organiser : ForensIQ (ULAB) in collaboration with Tech Global Institute

Project Description :

This project was centered around the ForensIQ workshop “Digital Forensics 101”, jointly organized by the University of Liberal Arts Bangladesh (ULAB) and Tech Global Institute. The session was designed to introduce students to the foundations of digital forensics and open-source investigations in Bangladesh, with a strong focus on hands-on learning. Participants explored the practical use of digital tools to trace locations using geo-data, verify audio authenticity, and detect AI-generated videos. The workshop highlighted how these investigative techniques can be applied to combat misinformation, strengthen media literacy, and enhance digital safety in the Bangladeshi context.

My Tasks and Responsibilities :

In this project, I was actively involved in supporting the smooth facilitation of the workshop. I assisted in receiving and welcoming the guest speakers and ensuring the event logistics ran seamlessly. During the hands-on sessions, I played a direct role in helping students apply the tools introduced, guiding them through practical exercises on geo-location tracing, audio verification, and identifying AI-generated media. By providing one-on-one assistance, I helped participants gain confidence in using these tools effectively. My contributions supported both the guest speakers and the student participants, ensuring that the workshop maintained its interactive and practice-based approach.







Learning Outcomes

Attending and assisting in the *Digital Forensics 101* workshop was a highly valuable experience that expanded both my technical and facilitation skills. By engaging with the tools demonstrated in the session, I gained **practical knowledge of geo-location tracing, audio verification, and AI-generated content detection**. These skills directly connected with my ongoing work on misinformation and strengthened my ability to apply digital forensic techniques in real investigative contexts.

Supporting students during the hands-on exercises also helped me develop **mentoring and peer-assistance skills**, as I guided them through the use of new tools and ensured they could successfully complete practical tasks. This experience reinforced my ability to **translate technical processes into simple, accessible steps** for others, an important skill for both research and outreach.

Finally, observing the collaborative environment of ForensIQ and Tech Global Institute emphasized the importance of **knowledge-sharing and interdisciplinary teamwork** in tackling digital challenges. The workshop not only deepened my technical expertise but also showed me how digital forensics can empower students and professionals alike to critically engage with information in today's digital world.

Project 8

Project Name: Mapping Extremist and Information Interference Channels

Date: 1 August 2025

Project Description:

This project focused on compiling and categorizing digital channels involved in **extremism, misinformation, and information interference**. The dataset systematically mapped social media pages, groups, and accounts across multiple categories, including Islamist extremism, Rohingya-linked extremism, Hindutva extremism, partisan political actors (BCL), suspected bot networks, and international interference campaigns. Each entry was documented with channel URLs, platform type, and content classification, and then assessed using the **DMET (Disinformation–Misinformation–Extremism Typology) framework** to determine levels of activity ranging from partisan narratives to violent extremism.

By analyzing these diverse networks, the project aimed to understand the **ecosystem of actors shaping online discourse in Bangladesh and the region**, as well as their cross-platform strategies. This work contributes to building a reference base that can inform further research, monitoring, and counter-extremism initiatives.

My Tasks and Responsibilities:

In this project, I was responsible for organizing and categorizing channel data from different sources into a structured dataset. I reviewed accounts and channels across Facebook, X (Twitter), YouTube, Telegram, and websites, and classified them by extremist affiliation (e.g., Islamist, Rohingya, Hindutva) or functional type (e.g., partisan actors, bots, interference networks). I also applied DMET classification to label levels of extremism, identified amplification vs. covert actors, and noted suspected automation patterns in bot accounts. Finally, I contributed to refining the dataset by ensuring accuracy, removing duplicates, and creating a clean reference list for analysis of extremist and interference actors.

Dataset : Mapping Extremist and Information Interference Channels

testing Channel List

File Edit View Insert Format Data Tools Extensions Help

Menus 100% 123 Default - 10 + B I U

M306

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
	Channel Platform	Channel Type	channel_name	channel_url	Content Type (IS/ AQIS/ Others)	Category based on content (Official/ Amplifier/ Covert)	Category: (VE/ Shared by VE/ Part of VE Network)	DMET_level_of stream	DMET_actor type	Primary country location of people who manage	Feb 2025 Status	Feb 2025 No.	Subscriptions/ Members	
2	Facebook	Facebook Page	Ramji Yusufur BH	https://www.facebook.com/pg/BlogRYS1/	AQIS	Amplifier	VE	Violent Extremis	Religious	Bangladesh (1),	Dormant		847	
3	Facebook	Facebook Page	Gazwatal Hind	https://www.facebook.com/1149172731915719/	Unaffiliated VE C	Covert	VE	Violent Extremis	Religious	Bangladesh (4)	Active		4,800	
4	Facebook	Facebook Page	Shaikh Tamin Al	https://www.facebook.com/sheikhtaminadlarnadi	AQIS	Official-Covert	VE	Terrorism	Religious	South Korea (2)	Dormant		12,000	
5	Facebook	Facebook Page	মুন্সফিকর	https://www.facebook.com/danufakimda/	Unaffiliated VE C	Covert	VE	Fringe Group	Religious	Bangladesh	Dormant		3,253	
6	Facebook	Facebook Page	Fidayi Hamla	https://www.facebook.com/fdayi.hamla/	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh	Active		2,000	
7	Facebook	Facebook Page	Karagar Theke A	https://www.facebook.com/karagarthekeshobani/	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh	Dormant		855	
8	Facebook	Facebook Page	Ummah Network	https://www.facebook.com/UmmahNetwork-307	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh	Dormant		4,400	
9	Facebook	Facebook Page	দ্বীন বিজয়ের ভান	https://www.facebook.com/%E0%A6%A6%E0%A6%	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh	Dormant		856	
10	Facebook	Facebook Page	Ibadah	https://www.facebook.com/ibadah-25418803160	AQIS	Amplifier	VE	Violent Extremis	Religious	Bangladesh (2)	Dormant		1,400	
11	Facebook	Facebook Page	শরণে নিকন	https://www.facebook.com/ClarificationDoubts/	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh, US	Dormant		755	
12	Facebook	Facebook Page	সালাহাচারিহ	https://www.facebook.com/salahacharia/	Unaffiliated VE C	Covert	VE	Terrorism	Religious	Bangladesh, UAI	Dormant		437	
13	Facebook	Facebook Page	জিরোয় শপ	https://www.facebook.com/%E0%A6%A6%E0%A6%	Unaffiliated VE C	Covert	VE	Terrorism	Religious	Zuwarah, Libya	Dormant		1,200	
14	Facebook	Facebook Profile	Saeed Adnan	https://www.facebook.com/profile.php?id=10002	AQIS	Covert	VE	Violent Extremis	Religious	Bangladesh	Dormant		682	
15	Facebook	Facebook Profile	Asif Adnan	https://www.facebook.com/AsifAdnanBH	AQIS	Covert	VE	Violent Extremis	Religious	Bangladesh	Active		334,000	
16	Facebook	Facebook Page	আব্দুল সাদেক রি	https://www.facebook.com/WAADAM24/	Unaffiliated VE C	Covert	VE	Terrorism	Religious	Bangladesh	Dormant		11,195	
17	Facebook	Facebook Page	উত্তমকান্ত হাসান	https://www.facebook.com/profile.php?id=10009	AQIS	Amplifier	VE	Terrorism	Religious	#N/A	Dormant		13,300	
18	Facebook	Facebook Page	An-Nashah	https://www.facebook.com/AnNashahMedia/	AQIS	Amplifier	VE	Terrorism	Religious	Bangladesh	Dormant		15,852	
19	Facebook	Facebook Page	Know Your Deen	https://www.facebook.com/KnowYourDeenBD/	AQIS	Covert	VE	Violent Extremis	Religious	Bangladesh (4)	Dormant		357	
20	Facebook	Facebook Page	আল কিতাব ইং	https://www.facebook.com/Al.kitah.medi/	AQIS	Amplifier	VE	Violent Extremis	Religious	Bangladesh	Active		1,200	
21	Facebook	Facebook Page	জাকার বায়েগুন	https://www.facebook.com/%E0%A6%A6%E0%A6%	AQIS	Amplifier	VE	Violent Extremis	Religious	Saudi Arabia (3)	Dormant		35,000	
22	Facebook	Facebook Page	লিংক মিডিয়া - Li	https://www.facebook.com/linkMedia2017/	AQIS	Amplifier	VE	Violent Extremis	Religious	Bangladesh	Dormant		38,000	

Islamist Extremist

Rohinova Extremist

Hindutva Extremist

RCL

Suspected Rot

Information Interference

Activate Windows
Go to Settings to activate Windows.

Findings

The mapping of extremist and information interference channels revealed a **diverse but interconnected ecosystem** of online actors. A large number of identified channels were concentrated in **Islamist extremist networks**, particularly on Facebook, with several pages amplifying narratives linked to AQIS and other violent extremist groups. These pages operated in mixed capacities—some as **official outlets**, others as **amplifiers or covert channels**—demonstrating a layered strategy of spreading content.

In parallel, the dataset showed the presence of **Rohingya extremist actors** active on Telegram, X (Twitter), YouTube, and dedicated websites. Many of these channels were tied to the Arakan Army or pro-Rohingya militant groups, often sharing materials that combined calls for autonomy with broader regional narratives. Their cross-platform presence highlighted the **transnational nature of Rohingya digital extremism**.

The **Hindutva extremist content** documented in the dataset was smaller in volume but notable for its cross-border impact. These Facebook and X accounts often framed narratives around religious identity and nationalism, aiming to **stoke tensions in Bangladesh's communal landscape**.

The project also identified partisan digital activity through a set of channels linked to the **Bangladesh Chhatra League (BCL)**. These accounts engaged in both political promotion and aggressive counter-narratives, positioning themselves as part of the larger information struggle in Bangladesh's online space.

Another critical finding was the detection of **suspected bot networks** on Facebook. These accounts displayed repetitive, automated posting behaviors, suggesting the use of algorithmic amplification to increase the reach of particular narratives. Such activity reflects a growing reliance on **coordinated inauthentic behavior** in Bangladesh's information ecosystem.

Finally, the **information interference category** revealed that Bangladeshi digital discourse is being shaped not only by domestic actors but also through **external influence operations**. Several identified accounts and channels were linked to international platforms and networks attempting to shape narratives around Bangladesh's politics and society.

Overall, the findings indicate that extremist and interference actors operate across ideological lines—**Islamist, Rohingya, Hindutva, partisan, and international**—but often converge in their use of **coordinated amplification strategies, bot networks, and cross-platform activity**. This underscores the complexity of Bangladesh's digital threat landscape and the need for ongoing monitoring and countermeasures.

Limitations and How I Overcome them :

Incomplete Metadata

- **Limitation:** Many channels lacked detailed metadata such as verified location, organizational background, or administrator identity, which made it difficult to confirm the authenticity of accounts.
- **Overcome:** I cross-checked posting times, language use, and linked accounts to infer probable locations and affiliations. In some cases, I compared these findings with open-source intelligence (OSINT) reports to validate assumptions.

Hidden or Ambiguous Affiliations

- **Limitation:** Several channels avoided explicit self-identification, making it unclear whether they were official outlets, amplifiers, or covert actors.

- **Overcome:** I analyzed posting style, thematic focus, and cross-platform activity to distinguish between **official**, **amplifier**, and **covert** roles, allowing more accurate categorization within the DMET framework.

Overlapping Narratives Across Categories

- **Limitation:** Some accounts disseminated mixed narratives—for example, combining extremist rhetoric with partisan politics or communal issues—leading to classification challenges.
- **Overcome:** I applied **multi-label categorization**, tagging accounts with multiple classifications where necessary, to preserve analytical nuance without oversimplification.

Detection of Automated Accounts

- **Limitation:** Identifying suspected bots was challenging due to the absence of direct technical metadata (e.g., posting frequency logs).
- **Overcome:** I relied on behavioral indicators such as repetitive posting, identical content across multiple accounts, and synchronized timing, using them as proxies for bot-like activity.

Dynamic and Evolving Networks

- **Limitation:** Several extremist and interference accounts were intermittently **dormant or deactivated**, creating gaps in longitudinal tracking.
- **Overcome:** I archived accessible data and noted activity status (active/dormant) in the dataset, ensuring that changes over time could be captured and reanalyzed in future updates.

